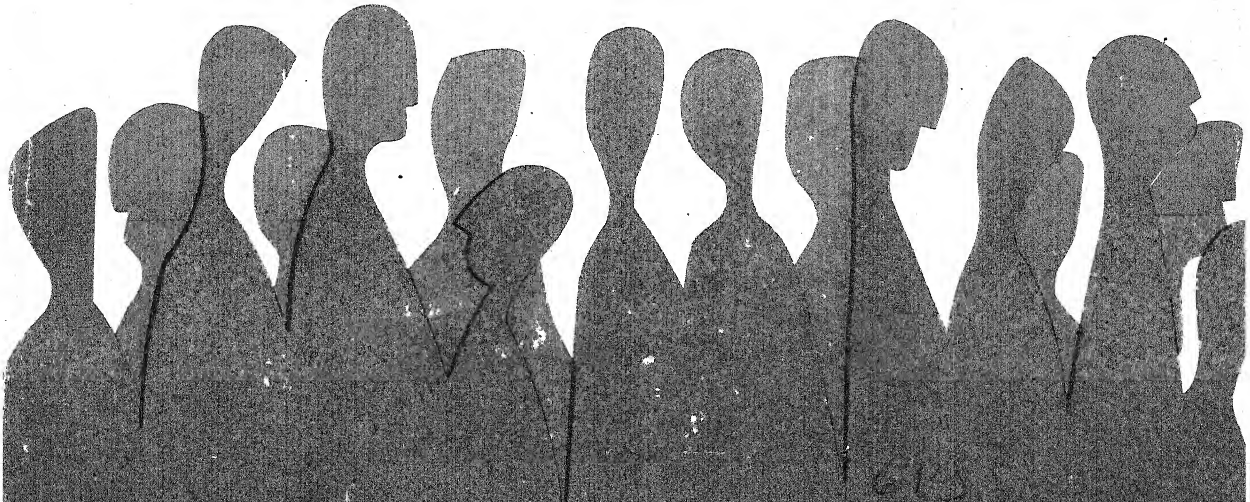


AN INTRODUCTION TO RESEARCH IN SOCIAL

CHOLOGY

CISES AND EXAMPLES

By Alan E. Gross, Barry E. Collins, James H. Bryan

Errata

An Introduction to Research in Social Psychology

Reproductions of the corrected errors are provided in order for you to copy the corrections directly in your book.

p. 108: top line: change columns *13* and *14* Data Sheet 2 to columns *2* and *3*.
"which caused the difference. Record responses on Data Sheet 2, columns ~~13~~ 2 and ~~14~~ 3."

p. 159: all columns labelled *Q3* should read *Q1*: all columns labelled *Q4* should read *Q2*.

Subject	Comparison				Positive				Negative			
	~~Q3~~	~~Q3~~^{2}	~~Q4~~	~~Q4~~^{2}	~~Q3~~	~~Q3~~^{2}	~~Q4~~	~~Q4~~^{2}	~~Q3~~	~~Q3~~^{2}	~~Q4~~	~~Q4~~^{2}
	Q1	Q1^2	Q2	Q2^2	Q1	Q1^2	Q2	Q2^2	Q1	Q1^2	Q2	Q2^2

p. 171: Third paragraph: The Patterson and Sechrest (1970) findings are more complex than indicated on this page. See reference on p. 175.
"people's attitudes toward others. In one such study, Patterson and Sechrest (1970) found that individuals who sat four feet away from another person were judged more friendly, extroverted, dominant and aggressive by that person than those who sat at distances of 2, 6, and 8 feet."

pp. 391, 395, 399, 403, 407: Do not use Box H. Revise Boxes I and L as shown.

G ☐ Add boxes A + B. — Total $N = (n_1 + n_2)$

Do not use Box H.

~~H ☐ Add Boxes C + E.~~ — ~~Total sum = ΣX~~

Square the number in Box C, square the number in Box E; add those two numbers and place the sum in Box I.

I ☐ ~~Square number in Box H.~~ — ~~$(\Sigma X)^2$~~

J ☐ Add Boxes D + F. — Total sum of squares = ΣX^2

K ☐ Multiply Box G by Box J. — (N) X (ΣX^2) or $N \Sigma X^2$

L ☐ Subtract ~~1~~ 2 from Box G. — N - ~~1~~ 2

pp. 412, 414, 416, 418: Revise Box M as shown.

M ☐ Multiply Box K by ~~Box N~~ A + B + C + D (the total number of Ss)

pp. 412, 414, 416, 418: Step 3: change *6* to *5* as shown.

"Step 3. If the χ^2 value in Box Q is larger than 3.84, these results would have occurred by chance less than ~~6~~ 5 times in 100 ($p < .05$)."

An Introduction to
RESEARCH IN
SOCIAL PSYCHOLOGY:
Exercises and Examples

An Introduction to RESEARCH IN SOCIAL PSYCHOLOGY: Exercises and Examples

ALAN E. GROSS
Ohio State University

BARRY E. COLLINS
University of Wisconsin

JAMES H. BRYAN
Northwestern University

John Wiley & Sons, Inc.
New York • London • Sydney • Toronto

Library of Congress Cataloging in Publication Data

Gross, Alan E 1936-
An introduction to research in social psychology.

Includes bibliographical references.
1. Social psychology—Research. 2. Social psychology—Problems, exercises, etc. I. Collins, Barry E., joint author. II. Bryan, James H., 1933- joint author. III. Title. IV. Title: Research in social psychology.

HM251.G7574 301.1'072 74-38967
ISBN 0-471-32815-4

Printed in the United States of America.

10 9 8 7 6 5 4 3 2 1

Contents

Preface

We wrote this book of exercises because we think that students in the typical undergraduate social psychology course are missing something. We know that even the most charismatic lecturer aided by an excellent text falls short of giving the student a thorough understanding of modern social psychological research. What is missing is experience—active participation in research design, data collection, and analysis. We believe that actual involvement with social psychology projects not only stimulates interest but facilitates mastery of substantive areas.

It may seem strange, then, that we begin a book whose major theme is "learning by doing" with a somewhat abstract Introduction dealing with methodology and philosophy of science. Although we feel that the value of traditional methodology chapters is limited when digested alone, we believe that abstract conceptual "bones" take on new meaning when they are filled in with the "meat" of concrete research experience. We are confident that the combination of abstract concept and concrete doing will enrich the learning experience. Thus the Introduction is intended to be used in two ways: it can be skimmed as a conceptual preview before the exercises are attempted, and it should be reread after some of the research techniques introduced here have become familiar. Hopefully, a second reading will help the student to place newly acquired research experience into a conceptual framework. The introductory chapter frequently refers to exercises that appear later in the book and, occasionally, when doing an exercise, the student is referred to the Introduction for fuller discussion of a relevant issue.

We have not however, limited the points of conceptual and methodological interest to the Introduction; in several exercises, interesting issues

and problems are discussed as they develop. Since we expect that most instructors will supplement this book with a standard text or several paperbacks, the introductory chapter is deliberately brief. Extensive discussions of methodology are available in Lindzey and Aronson's *The Handbook of Social Psychology Volume II,*[1] in Campbell and Stanley,[2] Kiesler, Collins, and Miller,[3] and Webb, Campbell, Sechrest, and Schwartz.[4]

Almost all of the exercises selected for this book have been tested in the classroom. We purposely omitted techniques that require elaborate laboratory equipment or sophisticated statistical procedures. Most of the necessary materials are provided in ready-to-use form. Simple statistical analyses are either explained in the text or are included in one of the appendixes. Each exercise is self-contained; detailed instructions, questionnaires, data recording forms, procedures for recording and analyzing data, and an annotated list of relevant background readings are included. Ten of the exercises (3 to 12) are followed by research reports that have been reprinted from professional journals. These articles were selected as examples of contemporary social psychological research because they are relevant to issues and methods discussed in the exercises and because they are written in a manner comprehensible to most undergraduates. Guidelines for preparing written research reports are given in Appendix E.

The exercises are arranged so that the student can learn such basic skills as interviewing, attitude scaling, unobtrusive measurement, and sampling before he is introduced to basic research paradigms in both laboratory (Exercises 4 to 6) and naturalistic (Exercises 7 to 9) settings. However, all of the exercises do not require the student to follow a cookbook formula. Frequently he is challenged to alter instructions or to devise additional data analyses. Instructors may wish to encourage such improvisation even when optional procedures are not suggested. Exercises 10 to 13 provide more advanced research experiences. Finally, in Exercise 14, the student is provided with a broad framework of current issues in the field of attitude change, and instead of being led through a series of procedures as in the earlier exercises, he is invited to select a problem area and to construct methods for answering questions within that area.

[1] G. Lindzey, and E. Aronson. *The handbook of social psychology.* Reading, Massachusetts: Addison-Wesley, 1968, II.

[2] D. T. Campbell, and J. C. Stanley. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching.* Chicago: Rand McNally, 1963.

[3] C. A. Kiesler, B. E. Collins, and N. Miller. *Attitude change.* New York: John Wiley, 1969, Chapter 2.

[4] E. J. Webb, D. T. Campbell, R. D. Schwartz, and L. Sechrest. *Unobtrusive measures.* Chicago: Rand McNally, 1966, Chapter 1.

Certain modifications or procedural changes may be desirable in various locations. For example, in sparsely populated schools or areas, each experimenter may be able to run only a very few subjects and it will be necessary to combine efforts in order to collect sufficient data. In some cases, forms are provided for such data pooling.

Occasionally an experimenter can become so absorbed in his work, that he may neglect his responsibilities to the human beings who serve as his subjects. It is imperative to safeguard the privacy, integrity, and feelings of the subjects who participate in the experiment. Especially in naturalistic settings or when the student decides to deviate from suggested procedures, *every precaution must be taken to safeguard the physical and mental welfare of those who participate,* and whenever possible, subjects should be fully informed about the research. A complete list of ethical guidelines[5] for the experimenter-subject relationship may be found in Appendix D.

Most of the procedures in this book have been pretested, but they are by no means guaranteed to produce results that consistently support given hypotheses. Despite the frustrations of messy reality (that is, the unlikelihood of obtaining a clear pattern of data from every exercise) we are hopeful that, through experience, the student will not only gain familiarity with existing techniques but that he will be challenged to adapt and apply his newly acquired scientific methodology to important problems in the social world around him.

We are grateful to Carol M. Bergdoll for her assistance in preparing the manuscript, and to James McMartin for his helpful suggestions and criticisms. We also would like to acknowledge the sporadic editorial comments of Annie Q. Ducque.

Alan E. Gross
Barry E. Collins
James H. Bryan

[5] American Psychological Association. Ethical standards of psychologists. *American psychologist.* The American Psychological Association, Inc., 1963, **XVIII**, 1, pp. 56-60. (For illustrations see American Psychological Association. *Casebook on ethical standards of psychologists.* The American Psychological Association, Inc., 1967).

An Introduction to
RESEARCH IN
SOCIAL PSYCHOLOGY:
Exercises and Examples

Introduction

Scientific Problem Solving in Social Psychology

In the late 1950s, shortly after the Soviet Union launched Sputnik, the United States committed substantial resources to a research program designed to investigate problems that had to be identified and solved before man could step on the moon. By 1960 considerable progress had been made, but NASA did not even know all the problems—much less the answers. Clearly the national confidence that an American would reach the moon before the end of 1970 did not stem from secure knowledge that the critical problems had been identified and solved.

But the United States taxpayers were not placing their dollars on a poor bet—although they knew that their scientists did not have all the answers, they had well-placed faith that the scientists knew *how* to find the answers. The physical and engineering sciences had a proven methodology for problem solving. If scientists had harnessed nuclear power, discovered antibiotics, built planes that fly faster than the speed of sound, and developed devices that receive pictures transmitted through airwaves, then these scientists could apply the same problem-solving methods to the obstacles that had to be overcome in order to place a man on the moon. It was faith in the American problem-solving capacity in general (and the *scientific method* in particular) which led to the heavy investment in the space program.

This book illustrates how some of the tactics, philosophies, and specific techniques from the scientific method that put a man on the moon can be applied also to problems in social psychology. For instance, on March 14, 1964, New York citizens watched more than 30 minutes as an assailant slowly killed Kitty Genovese. Not a single one of the onlookers intervened even to the extent of placing a telephone call to the police. Why? Could

changes in our laws, our child-rearing practices, our public school curricula, or the design of our urban environment increase the likelihood that American citizens will help each other?

Although these are questions that well deserve the attention of the philosopher, theologian, educator, parent, citizen, and legislator, we believe that the scientific social psychologist can make important contributions to problem-solving activities stimulated by this incident. In fact, a few social psychologists (for example, Latané & Darley, 1970) have recently become involved in scientific research on "helping behavior." Exercise 9 illustrates a scientific approach to one aspect of this problem.

Some exercises in this book are representative of more traditional research, and no topic has been more central in the social psychology of this century than group process. Why do group discussions on vital topics—whether international diplomatic discussions, foreman-worker conferences, student council deliberations, or family discussions—often degenerate into useless interpersonal bickering? What can we learn about the dynamics of group interaction that would aid a group to function effectively without dehumanizing group members into organization men, men in grey flannel suits, or unthinking automatons? Exercise 11 illustrates some group research problems and techniques.

What do we know about attitudes and opinions? Can they be measured? (See Exercise 2.) How are they formed? How are they changed? (See Exercises 5, 6, and 14.) What are the laws that determine how one person perceives another? (See Exercise 13.) How do human beings use spatial arrangement to express a social message? Do people stand closer to friends than to enemies? (See Exercise 7.)

Textbooks and journals on social psychology suggest a number of provocative answers to questions like these. Our focus in this text, however, is not on *what* we know about these kinds of social problems. Instead, this is an introductory laboratory manual on the research techniques that social psychologists have used to identify and answer questions of this sort. Thus we shall emphasize *how* scientific social psychologists have attacked these problems. What is their philosophy of science? What are some of the strategies and techniques they have used in their research efforts? What are some of the specific methodological techniques that they have found useful in their search for answers to social problems?

The essence of the scientific method is that it is a *public* problem-solving strategy. The procedure for producing scientific results must be public because other investigators must understand the research procedures of their fellow scientists. When a procedure is accurately communicated, other scientists should be able to reproduce the procedures and obtain the same results in their own laboratories. Much of the social scientist's concern with quanti-

fication and precise measurement, for instance, stems from the fact that quantification and precise measurement makes accurate, public communication easier.

In this manual we shall provide an introduction to some of the scientific research strategies that scientific social psychologists have found useful in the analysis of social issues. We shall illustrate how these strategies can be applied to the understanding of some important social phenomena. The philosophy of science, general orientation, and specific methodological procedures illustrated in this manual are among the most promising strategies for gaining insight into how human beings relate to each other.

STEPS IN FORMULATING A RESEARCH DESIGN

Social psychology certainly does not suffer from a shortage of interesting problems! How are we to understand the violence of one human being toward another—whether it be a youth gang smashing a window, a middle-aged conspiracy to embezzle funds, a gangland-style killing, or two nations engaged in aerial warfare? Where does one start to gain insight into bigoted attitudes that one geographic, ethnic, or religious group holds toward another? What are the dynamics of bitter competition between teams? Conversely, what are the dynamics that lead to intimate cooperation among the members of a football team? Would changes in the typical American family structure and its current practices of child rearing produce significant changes in adult personality? The first step for the social scientist, then, is to formulate a researchable question. As always the question asked and the form in which it is asked strongly influences the probable answers.

1. Select a Problem Area, Situation, or Phenomena

When a social scientist formulates a question he should have two criteria in mind: (1) relevance and importance, and (2) researchability.

Relevance and Importance. If a scientist should be successful in finding an answer to the question he poses, will anyone be interested? The number of questions on which the social scientist might collect data is almost unlimited. Thus he has an obligation to choose a question that needs answering. Will the answer to this particular question increase our basic theoretical understanding of human behavior? Will the answer to this particular question lead to any practical applications? Will the answer to this particular question test a more general theory? In other words, is the question "relevant" either to a general social psychological theory or to the analysis of a particular applied problem?

Researchability. For the researcher who is committed to the scientific method, it is not sufficient that his questions be interesting ones. A second criterion is necessary: Are the technology, methodology, and financial and human resources available to provide a meaningful answer? No matter how theoretically, philosophically, ethically, and practically interesting is his question, a social scientist wastes his time when he works on questions for which there is no adequate problem-solving methodology available.

Thus, a good social scientist must be versed in research methodology and statistics *at the time he formulates his research question*. He must anticipate possible ethical problems, problems in data analysis, the availability of appropriate measurement devices, the availability of appropriate sites in which to test his hypothesis, and the availability of the right kind of subjects for the experiment. He must be aware of possible alternative explanations which could make his answer so ambiguous that it is less useful than no answer at all. All of these factors should influence both problem selection and research design.

Altogether too frequently, a social scientist finds himself in this dilemma: Sophisticated methodological procedures, which are most likely to provide unequivocal, useful, and relevant answers may not be applicable to the most interesting questions. And, when he finds an interesting question, he may be unable to find an adequate methodology or to marshal the necessary resources to provide a meaningful answer. Thus the selection of a research question is almost inevitably the result of some kind of compromise. The scientist must settle for the most interesting question for which there is an adequate methodology to find an unequivocal and meaningful answer.

2. Formulate a Specific If-Then Hypothesis

Philosophers of science have not yet agreed on a precise statement which specifies the appropriate goals and procedures of scientific analysis. But, without too much danger of oversimplification, we can state that a major goal of social science is to understand social phenomena. Whether he is interested in the indifference of bystanders observing a crisis, interpersonal bickering at a conference, or violent mob behavior, the social psychologist searches for the *how* and *whys* of the phenomena he is studying. His aim is to analyze the phenomena into their constituent elements, variables, or components; then his aim is to state the relationships among these elements or variables.

This last sentence can be restated in a number of different ways. For instance, we could say one aim of science is *prediction*. When the social scientist is interested in prediction, he uses the information that he has about one variable to predict something about another variable about which he has less information. For our purposes we can refer to the blank spaces in the

"If . . . then . . ." question as variables. The if-then format can be stated in the form, "If (a large or small amount) of *variable 1* is present, then (a large or small amount) of *variable 2* will also be present." For example, the scientist may know a great deal about a student's high school grades, his objective test scores, his performance on extracurricular activities; but he may know relatively little about how that student will perform in college. If he is interested in the unknown variable(s), he may conduct a scientific study to discover the relationship between those variables about which he has some information and those variables about which he has little or no information. The basic form of his question can usually be stated in an if-then hypothesis format. For instance, "If a student has *good high school grades* then he is likely to *graduate from college*."

Other scientists are less concerned with the immediate practical problem of prediction and are more interested in *understanding* and *explaining* a social phenomenon. For example, how should the scientist go about "explaining" the process by which an individual evolves toward a position of leadership in society? The if-then hypothesis format proves useful when the scientist seeks explanation as well as when he seeks prediction. For example, the following statements may contribute to our understanding and explanation of the leadership selection process. If the student's *father was a leader*, then he will be *selected as a leader.* If a student indicates *unusual creativity* in his classwork, then he will be *selected as a leader.* If a student *conforms to the norms and expectations of his peers,* then he will be *selected as a leader.* If a student *attends an ivy league school,* then he will be *selected as a leader.* If a student *takes a speed reading course,* then he will be *selected as a leader.*

These examples are typical of the empirical, data-oriented style of explanation. Other scientists find the if-then hypothesis format useful for *theory testing*. Whether the theory is a few paragraphs of homespun philosophy, an elaborate set of logical statements in a formal logical theory, a precise mathematical statement, or a middle range statement such as, "As group size increases, each individual group member will feel less responsibility," the theory cannot be tested until it is translated into if-then propositions. If decisions are *made in a group* (compared to individual decision making), then the decisions will be *more risky* (as measured by a pencil and paper test). If group *size increases*, then individuals will feel *less responsibility* (be less likely to report smoke pumped into the air conditioning by the experimenter).

Perhaps most importantly, the if-then format is a useful empirical analysis of *causation*. If the scientist wants to know what causes a variable such as altruistic or helping behavior, the most useful procedure is to look for the antecedent conditions. If we know what occurs and what does not occur

immediately prior to an altruistic act, we gain insight into what causes and what does not cause the altruistic act. Thus, if we find that the following statement: "If the bystander *foresees a personal danger in helping*, then he is *unlikely to offer help*," is true, then we may be willing to conclude that perception of personal risk causes selfish or unaltruistic behavior.

Let us consider another example in which the leap from the empirical if-then statement to causal theory is a little larger. Suppose it is the fact that, "If a *large number* of bystanders are witnessing an emergency, then any single bystander is *unlikely to help*." (See Exercise 9.) From this if-then empirical hypothesis, we may be willing to make a more theoretical statement: "A diffusion of responsibility inhibits helping behavior." A theoretical statement should be supported by several if-then statements. Therefore, the following if-then statement would considerably increase our confidence in our diffusion of responsibility explanation: "If an individual perceives that others in the group *are more capable of helping* than he, then he is *unlikely to help*." Thus, individual if-then hypotheses may be viewed as the bricks of theory building; that is, a number of different findings in support of a theoretical statement give us confidence in that particular explanation for a class of social phenomena.

In summary, the first step in the formulation of a scientific research question is to select a problem area, situation, or phenomena. Once this is done, the scientist must decide the general level of analysis—whether concrete or abstract—at which he wishes to work. Once he has accomplished this, the next step is to formulate his scientific question in the empirically testable if-then question format.

3. Construct a Literary Definition for Both Variables in the If-Then Hypothesis

As was previously stated, an essential characteristic of the scientific method is that it be public. Most of the rules, procedures, and philosophies associated with the scientific method are designed to insure that one scientist can understand what another is doing and thinking and to maximize the possibility that a discovery or insight reported by one scientist can be reproduced by another scientist. Thus the first and most elementary job of the scientist is to construct definitions of the variables he is studying which make his work public by satisfactorily communicating what he is doing to the rest of the scientific community.

A *literary definition* of a variable *uses the everyday spoken language of the scientific community to help one scientist communicate his general conceptions of a variable to another scientist.* A literary definition may be rich with example and metaphor; and it is not always precise. A literary

definition is a set of words from the everyday language which relies on consensus among users for its meaning. Dictionary definitions often formalize much of this user consensus. A literary definition relies on prose, metaphor, illustration, and rebuttal in order to communicate the meaning from the researcher to his intended audience. Literary definitions are typically rich in content and broad in scope; it is an unusual literary definition that can be transformed into a measurement procedure with equal richness and breadth.

The precision of such definitions can be increased considerably when the scientist supplies specific operational definitions. *An operational definition describes in detail the specific set of techniques for measuring the variable that is being defined.* Communications using operational definitions are often dry, lengthy, and literal; and they are relatively precise and accurate.

Some philosophers of science have denied the usefulness of literary definitions. They say, for instance, that intelligence *is* the score from a particular intelligence test—literally that and *nothing* else. But a single measurement procedure or operational definition falls far short of what most people mean by intelligence. Any particular measurement procedure is only one of a set of many measurement procedures which most of us would be willing to lump together under the general label "intelligence." We may be willing to rely on our intuition and say that successful performance in an English class and successful performance in a history class are both likely to reflect a broader concept which we call intelligence. Or we may know that a number of different measurement procedures are empirically intercorrelated. When a set of variables are highly intercorrelated, many of us may be willing to infer that each operational definition is simply another way of measuring the same variable or theoretical concept.

Any particular operational definition (measurement procedure) is only one of a set of many measurement procedures that can be derived from a good literary definition of a variable. In a particular experiment on helping, for instance, the scientist purposely limits his attention to a single measurement procedure such as a subway rider's willingness to contribute a dime for a phone call or a college student's willingness to report smoke coming out of the ventilation system. But it is clear that the experimenter's own conception of "helping" includes more than whether the accomplice gets a dime or whether the subject reports smoke to the secretary. Similarly, a number of investigators have studied aggression by observing the intensity of electric shock one subject is willing to deliver to another or the number of times a child punches a bobo doll; but it seems unlikely that either or both of these operational definitions capture all of what these investigators have in mind when they use the word "aggression." Thus most scientific social scientists present a literary definition for their concepts.

4. Choose a Method to Quantify and Measure the Two Variables in the If-Then Hypothesis

After selecting a variable and constructing a literary definition of it, the next step in the formulation of a research question is to specify an operational definition. An operational definition is the specific procedure for creating or measuring the variables being studied. The scientist studying group phenomena may be interested in a variable such as group cohesiveness. As his literary definition, he may state that group cohesiveness is the sum of all the forces that tend to keep a group member inside the group. For his operational definition, he may compute a number based on the degree of liking each group member expresses for the other members of the group on a questionnaire. Once he has accepted this operational definition, the scientist has moved from the verbal world of theory to the empirical world of hypothesis testing. His statement, "If a group is threatened by an outside agent, then it will become more cohesive," is not directly testable. But the statement, "If a neighboring juvenile gang makes incursions on the group's own territory (an operational definition of external threat), then the threatened gang members will indicate greater liking for each other on a pencil and paper test (an operational definition of group cohesiveness)," is one that can be tested empirically.

In summary, after the scientist has selected a problem area or situation, decided on his level of analysis, formulated a specific if-then hypothesis, and constructed a conceptual or a literary definition of his two variables, he must then construct an operational definition which specifies exactly how the two variables are to be quantified and measured.

An operational definition serves two essential functions. First, it meets the minimum essential qualification for science—that it be public. Operational definitions specify particular measurement procedures and use a language about which there is an extremely high degree of consensus among the members of the scientific community. The operational definition insures that one scientist will be able to replicate a study done by a fellow scientist. If two scientists are studying cooperation, they may not agree on a conceptual definition; but they should be able to understand and communicate precisely on their operational definitions of cooperation. Second, operational definitions bridge the gap between theory and empirical tests. When the scientist constructs an operational definition to correspond with a literary or conceptual definition, he has specified the empirical measurement procedures that can be used to test the theoretical and conceptual notions contained in the literary definition.

Multiple Operationalism. A single operational definition is, almost inevitably, narrower and less rich than the literary definition. But does that mean

that empirical science must always be "constrained," "trivial," "literal," and less "relevant" than philosophy and theory? Although a scientist will seldom, if ever, capture the full theoretical richness of his concept in any *single* operational definition, the conscientious psychologist will construct several operational definitions for each of his literary definitions. Although no single operational definition catches the entire essence of his literary concept, a series of operational definitions and measurement procedures in concert may do just that.

For example, it is doubtful that many of us feel that "the intensity of shock a subject is willing to deliver to another subject" captures completely what we mean by hostility. But a longer list of operational definitions including derogatory verbal comments, refusing to return lost articles to the owner, and awarding a less than equal division of some scarce resource—taken together—do represent a reasonable empirical approximation of our literary concept of hostility.

Ideally, a literary definition should suggest a wide variety of operational definitions and measurement procedures. A well-constructed literary definition should allow a number of scientists to independently construct several operational definitions, each of which is a reasonable operational definition for that particular concept.

In summary, a scientist should not be satisfied once he has tested a theoretical proposition with a single operational definition. Our confidence in the validity of his conceptual or literary if-then statement is increased considerably when the proposition is tested with a number of different operational definitions. Ideally, the operational definitions should be as methodologically different as is possible—having in common only the fact that the scientific community agrees that each operational definition refers to the same literary or conceptual variable.

5. Select a Research or Question-Answering Strategy

With his problem area selected, the level of analysis decided, the specific if-then hypothesis formulated, a conceptual (literary) definition constructed for each of the variables, an operational definition or measurement procedure selected for each of the two variables, the next step is to select an appropriate strategy for making an empirical test. There are scores of considerations that could be subsumed under the heading of "research strategy"; but the discussion in this section will be limited to the comparison of *correlational* and *experimental* methods. In the correlational method, the experimenter does not interfere with the phenomena he is studying. He only measures the two variables in his if-then research hypothesis. To illustrate, consider the following conceptual if-then statement: "If an organism has

adequate diet during infancy, performance during adult years will be superior." If the investigator chose the correlational approach, he would look for a number of people who differed in the degree to which they had an adequate diet during infancy. He would then measure both of his variables (adequacy of diet in infancy and performance in adult years) on a moderate number of such cases. He would next inspect his data to discover whether instances of good diets were associated with high adult performance and instances of poor diets were associated with low adult performance. He might select a number of individuals and actually observe their diet during infancy, store his data, and then measure the adequacy of performance during adult years. Or he might divide the world into a number of areas that differ widely in the adequacy of infant diets and subsequently quantify the quality of infant diet in each geographical area and also quantify the quality of adult performance. Any independent unit of observation is acceptable. He might quantify infancy diet and adult performance in a variety of ethnic groups, members of different occupations, or different countries. Or he might compare infants raised in years where little food was available with infants raised in years where food was plentiful. He could work with humans, plants, or animals. In each case, he can test his if-then proposition by simply quantifying both of the variables and observing whether the two variables are, in fact, empirically associated.

In the experimental method (as contrasted with the correlational method), the experimenter actively intervenes in the process he is studying. He randomly assigns some subject to rich diets and other subjects to poor diets, while observing (again using a statistical index) whether the groups that he treated differently on variable 1 of the if-then statement differ on variable 2.

The experimental-random assignment strategy is superior from a scientific point of view because it provides better information about whether variable 1 *causes* variable 2 in the if-then hypothesis. However, the experimental-random assignment procedure also requires that the experimenter intervene actively, which may be physically, financially, or ethically impractical. Within our present technology, for instance, it seems doubtful that an experimenter could randomly assign some years to heavy rainfall (and thus a good diet in infancy) and other years to low rainfall. Even if technologically possible, the investigator may have ethical considerations that prevent him from intervening to that extent into the lives of his subjects. Thus, the constraints of the experimental-random assignment method, namely that the experimenter must actively intervene in the lives of his subjects, may force the experimenter to test his hypothesis in a less rigorous form, on a less relevant population, or with a narrower range of variation in his two variables. Although the experimental method does provide superior scientific and causal information, the experimenter may well choose the correlational

technique if the restraints of the experimental procedures prevent him from testing his hypothesis in the form that interests him most.

CORRELATIONAL ANALYSIS

The distinction between experimental and correlational research can be clarified by a few specific examples. To begin with, let us consider the conceptual hypothesis, "If a group experiences a threat from outside the group, they are more likely to cooperate among themselves within the group." If the scientist has decided to work at a global level of analysis, he might operationalize variable 1 (external threat) as the "periods of time when a country was at war" and variable 2 (cooperation) as "degree of internal political strife." Suppose that, in fact, his analysis revealed that the highest ratings by historians for internal strife were obtained in peacetime and that low levels of internal political dissension occurred during wartime. If this pattern of data had been obtained from a representative sample of countries and had been submitted to appropriate statistical analysis, we could conclude that external threat and cooperation (as operationally defined) were related to each other. With this example as an illustration, the next sections return to a more general discussion of correlational analysis.

Language of Correlational Analysis. If a high level of one variable is consistently associated with a high level of another variable, we say that these two variables are *positively* correlated. If high levels of one variable are associated with low levels of another variable we say that they are *negatively* correlated. In either case, a high degree of association or correlation gives us a basis for *predicting* from the known value of one variable to a second unknown variable. In other words, a correlation is a descriptive statement about the world; it says that a high level of one variable is positively related, unrelated, or negatively related to a second variable.

For example, height and weight are usually positively correlated; therefore, if we know that a particular individual is taller than average, we can predict that he will also weigh heavier than average.

Correlation and Causation. Although we accept the descriptive fact that height and weight "go together" and although we can *predict* a person's weight with some accuracy if we know his height, we do not believe that his height *caused* his weight, nor that his weight *caused* his height. Similarly, the correlation between external threat and cooperation does not demonstrate that external threat *causes* cooperation.

When a change in one variable is correlated with a change in another variable, it is tempting to assume that a cause-effect relationship exists; and such a relationship may indeed exist. But the descriptive correlation *always*

leaves room for several causal interpretations. For example, cooperative citizens might so threaten neighboring countries that they start preventive wars; thus the descriptive and predictive correlation between external threat and internal cooperation could mean that internal cooperation causes external threats.

In general a correlation can have three separate kinds of explanations: (a) variable 1 may cause variable 2, (b) variable 2 may cause variable 1, or (c) both may result from some third common cause—variable 3 might cause both 1 and 2. The weight-height correlation illustrates the "third common cause" explanation of correlation. We presume that some combination of genetic and environmental variables cause both his height and weight. Large parents (variable 3) produce tall (variable 1) and heavy (variable 2) children. The fact that height and weight are correlated is explained by the fact that parental size causes both height and weight in children.

Thus correlation does not *necessarily* establish causation; but the correlational analyses can often provide some causal information. Experience or logic allows us to eliminate alternative explanations for correlations such that we are left with only one plausible causal interpretation. For example, if we find that the length of a boy's trousers are correlated with his height, three kinds of explanations are possible: trouser length leads to tallness, a boy's height may cause his trousers to be long, or his mother's ambitions may cause her to buy oversized trousers and also to stimulate the boy to grow taller. Since only one of these explanations is particularly plausible, we may be willing to conclude tentatively that a boy's height causes his trouser length to be long. Similarly, if age is correlated with income, we may be willing to rule out the income causes age interpretation on logical grounds.

Now let us consider an example in which correlation gives only minimal information about causality; the often-reported positive correlation between socioeconomic status and intelligence. It is well documented that children from families with high levels of education and income score higher on standard IQ tests than do children whose parents have relatively little education and earnings. What does this fact mean? In this case, persuasive arguments can be made for each of the three alternative causal interpretations of this relationship:

1. Socioeconomic status or class causes intelligence. The low-income family may provide a stereotyped or unstable environment that directly impairs the full development of intelligence. Or higher income families may provide an enriched, intellectual or dietary environment in which superior intelligence can develop.

2. Intelligence causes social class. People with high IQs may achieve high levels of education and, as a consequence, bring home larger paychecks.

3. Some third variable causes both social class and intelligence. Society discriminates against ethnic groups so that they are unable to achieve either the schooling necessary to score well on measures of intelligence or to secure the opportunities needed to achieve high levels of education and income. Discrimination in favor of some groups and against others would cause certain individuals to be both smart and well paid, while causing others to score poorly on standard tests and to bring home a relatively small paycheck.

We have listed only a few general explanations for the relationship between social class and intelligence. The reader should be able to find several reasons why intelligence might lead to higher social class, other reasons people with high social class become intelligent, and still more possibilities for third variables that might affect both intelligence and social class. Motivational variables, type of diet, even physical appearance might be invoked as possible third variables.

The point is that when we are limited to correlational studies, it is extremely difficult to pin down the underlying causal connections. A correlation cannot be interpreted unequivocally to demonstrate a single causal relationship between the two variables in the if-then research questions.

EXPERIMENTAL ANALYSIS

The psychologist who wants strong causal information in answer to his if-then question will probably use the experimental question-answering strategy. In an experiment, the scientist deliberately attempts to induce change in the behavior of his subjects. If a second variable is altered as the result of the artificially induced change in the first, we say that variable 1 must cause variable 2. In order to conduct an experiment, the investigator *randomly* assigns his subjects to at least two groups. A random method of assignment assures that each subject has a probability equal to that of any other subject of being included in any of the experimental groups. One subject does not have a better chance of being assigned to one group than to another group. Thus, at the point of random assignment, it is a good statistical bet that the two groups are equal in all respects. If the experimenter treats the two groups differently and a posttest reveals a difference, the experimenter's treatment or manipulation probably caused the measured difference on the posttest. For example, an experimenter might randomly assign football players to two groups of teams. The random assignment makes it likely that the group of teams are equal at the point of random assignment. One group of teams would be exposed to external threat (competition with another team, a threat to cut off funds for football, and so on). The other group of teams would not be threatened. After a time, the experimenter could measure the cooperation among team members. If the

threatened teams are more cooperative, it seems likely that the threat *caused* the cooperation. The random assignment makes it likely that the groups were equal when they were assigned, so the observed difference probably was caused by something that happened to the teams after they were randomly divided into two groups.

Independent and Dependent Variables. In the experimental analysis, the general if-then question is often stated in terms of independent and dependent variables: If the experimenter manipulates or actively intervenes to *produce a change* in the *independent* variable, then he will be able to *measure* changes in the *dependent* variable. In the preceding example about football teams, external threat is a manipulated or independent variable and intrasquad cooperation is the observed or dependent variable. The scientist who wants strong causal information must select and isolate at least one variable to manipulate (the independent variable) and at least one way to measure or observe the effects of his manipulations (the dependent variable). Since the investigator does not intervene actively in correlational analysis, the independent versus dependent distinction is not made between the variables of an if-then statement tested with correlational analyses.

Continuous and Dichotomous Variables. Some dependent variables are *dichotomous*, that is, the behavior is recorded as either occurring or not occurring. Donating or not donating to charity is a dichotomous dependent variable (Exercise 8); or the dependent variables may also be measured along a continuous dimension; for example, an experimenter can observe the amount of money contributed to charity, the degree to which one cheats at cards, the amount of attitude change, and the intensity of cooperation or conflict.

As with dependent variables, the experimenter may choose dichotomous independent variables. A dichotomous independent variable is usually of the type where a specific situation, stimulus, or condition is present in one treatment but absent in another—for example, one group may role play, a second group does not. Of course the experimenter can decide to manipulate an independent variable along a dimension. For example, some subjects might role play for one minute, others for four minutes, and still others for ten. Or, if the experimenter is interested in how the size of a group affects amount of conversation and eye contact, he can manipulate his independent variable—group size—by systematically changing the number of people in his groups.

Summary

There are five steps in selecting a research design: (1) Select a problem area, situation, or phenomena according to two criteria: (a) relevance and

importance and (b) researchability. (2) Formulate a specific if-then hypothesis. Construct (3) literary and (4) operational definitions for both variables in the if-then hypothesis. (5) Select an appropriate correlational or experimental research strategy.

CHANCE AS AN ALTERNATIVE EXPLANATION FOR OBSERVED DIFFERENCES

Many students argue, "random assignment does not assure exact equivalence. Just by chance, it will almost always happen that groups of randomly assigned subjects are not exactly alike." That is correct. If subjects are randomly divided into two groups, they will probably differ on the dependent measure even if no independent variable or treatment is introduced. Usually there will be *some* difference between two groups even when subjects are randomly assigned. If an experimenter induces different levels of his independent variable (for example, external threat) and observes differences in his dependent variable (for example, amount of cooperation) he must have some way of comparing whatever differences he observes with differences that might have been expected on the basis of chance alone.

In other words, there are always two possible explanations for a difference between two randomly assigned groups that have received differential experimental treatments: (1) the observed difference might have been caused by the independent variable or (2) the differences might be because of "chance" initial differences in the subjects.

How is the researcher to know how large a difference might have been expected on the basis of chance alone? Fortunately for the experimental social psychologist, there are mathematical models that answer this question. These models specify exactly how often any particular observed difference would have occurred by chance when the groups were randomly assigned. In other words, an experimenter can compute the difference between the experimental groups that he actually observes. Then he uses statistical analyses to tell him how often that difference would have occurred by chance alone. Of course, even the largest difference might occur by chance once in a million or once in a hundred-million times. Thus the scientist never completely discounts chance as an alternative explanation for an observed difference; but statistical analyses can tell us if chance is a very likely alternative explanation for the observed difference.

Most social scientists have adopted an arbitrary rule of thumb. They have decided that they will not take any observed difference seriously if it could have been produced by chance more often than five times in a hundred. If statistical analysis indicates that an observed difference would have been produced by chance less often than five times in a hundred, we say that it is

significant at the .05 level ($p < .05$). If statistical analysis reveals that a difference would have been produced by chance fewer than one in one hundred times, we say that it is significant at the .01 level ($p < .01$).

Computation of the statistics as outlined in the Appendices may seem time consuming and difficult and, as those of you who have taken a statistics course know, the theoretical and mathematical assumptions used in deriving computational procedures are complex. But these computational and theoretical complexities should not obscure the basic purpose of a statistical test. The purpose is simply to obtain a number that tells us how often our observed difference could have occurred by chance. If it is unlikely that our observed difference was produced by chance, then we shall conclude that the observed difference was caused by our independent manipulation.

Matching as an Unsatisfactory Alternative to Random Assignment

Many students become skeptical when they are told that random assignment to different treatment groups of conditions is a prerequisite for making sound causal inferences. They wonder: "How does random assignment insure the equality of two groups? If I randomly assigned all of my friends into two groups, I know perfectly well that they will not be exactly equal. The average hostility of one group will be somewhat different from the hostility of the other, one group will have a larger percentage of blue eyes, one group will be slightly smarter than the other, and my friendship will be somewhat more intimate with one or the other group." Many students suggest that the best method for producing equal groups at the start of an experiment is to measure all subjects on all of the important variables that one can think of and then, based on these measurements, the experimenter can create two groups of subjects that are exactly equal. This intuitively appealing procedure is termed *matching*.

As appealing as it might seem, there are several serious problems with matching. A perseverant and careful scientist might attempt to measure his subjects on all of the dimensions that he believes are important, for example, intelligence, level of motivation, and class. Then he might divide these premeasured subjects such that they form two exactly equal groups. But how do we know that they do not differ in some other respect? Perhaps one set of groups has more friends in it than another set of groups. It is virtually impossible to match subjects on every conceivable relevant variable. Even if we were to match subjects on 1000 variables, we have no way of knowing that the two groups are not different on variable 1001. A fundamental criticism of matching is that the *matching does not assure us that the groups are completely equivalent.*

Second, matching is frequently impractical. Subjects may not be available for lengthy classification procedures or certain relevant characteristics may be unknown or inaccessible. A third criticism of matching is that the

measurement or testing process necessary for matching may affect how the subjects behave during the experiment. Such testing or observation may *sensitize* subjects to the research hypothesis and enhance "guinea-pig" effects (Webb, Campbell, Schwartz, & Sechrest, 1966), that is, if subjects are aware of the experimenter's purpose they may be unwilling or unable to respond "naturally" (Orne, 1962).

When a given subject characteristic is very likely to affect experimental results, it may be desirable to match subjects or groups of subjects and subsequently to assign them randomly to experimental conditions. Such a procedure combines advantages of both methods. For example, in an experiment designed to determine what kinds of messages are most effective in changing racial attitudes, it would be advantageous to consider the differential effects of the various messages on racial bigots, moderates, and liberals. A group of each kind of subject could be exposed to one of the persuasive communications and attitudes would later be measured. In order to combine matching and random assignment, in this case, the experimenter could divide his subjects into approximately equal groups of racial bigots, racial moderates, and racial liberals based on racial attitude test scores. Within each of these three groups, subjects could be randomly assigned to experimental conditions. Thus the experimenter would be assured that each and every experimental condition was made up of approximately one-third racial bigots, one-third racial moderates, and one-third racial liberals. Furthermore, he would have a true experiment because his subjects were assigned to conditions randomly, that is, each subject had an equal chance of being in any experimental condition.

Summary. Since matching has several serious faults and statistics allow us to evaluate chance as an alternative explanation of observed differences, random assignment is the best way to insure that groups are equal at the beginning of an experiment. If two groups are significantly different by a statistical test, we infer that something happened to them after random assignment that *caused* them to be different. If the manipulation creating the independent variable is the only thing different that has happened to one experimental group versus another, then we infer that the independent variable caused an observed difference in the dependent variable.

OBSTACLES TO UNAMBIGUOUS INTERPRETATION OF EXPERIMENTAL RESULTS

1. Complex or Confounding Independent Manipulations

We say a manipulation is confounded when the operational definition of the independent variable is composed of several components or dimensions —any one of which could have produced the observed effects on the

dependent variables. Kornhauser (1930) reported that students in his economics class at the University of Chicago became more liberal. But, suppose we were to find that the factual information in Kornhauser's textbook was, in and of itself, sufficient to increase economic liberalism. Furthermore, suppose that Kornhauser's own liberal political philosophy could, in and of itself, produce an increase in economic liberalism. If this were the case, we would say that Kornhauser's original experiment was confounded. Because each of these two variables (textbook information and lecturer attitudes) were mixed together in the same operational definition ("Kornhauser's economics course"). The original experiment cannot tell us which of these two variables was actually causing the increase in economic liberalism.

Although we have no way of knowing for sure, Kornhauser probably wanted to test the hypothesis that the information he transmitted in his lecture combined with material assimilated from the textbook produced a change in economic attitudes. If this was the hypothesis he wanted to test, his experiment was inadequate. He confounded factual lecture and book information with expression of his own values in the same experimental manipulation or independent variable. Therefore, he cannot conclude that transmission of facts caused attitude change; factual information and expression of instructor's attitudes are inextricably combined in the same independent manipulation. It is impossible to know which aspects of his complex manipulation (the entire course) were actually responsible for the observed changes.

The implication of this discussion of confounding is that the treatment of various experimental groups should differ in as few respects as possible. The fewer differences between two experimental conditions, the more certain we are of just exactly what caused the observed differences in the dependent variable. The more refined the manipulation of a particular experiment, the further along we are on the analytic trail that leads to understanding human behavior.

In addition to complex manipulations that are specific to a single experiment, there are a number of more general confounds that plague interpretation of data. These general confounds are seldom intended to become part of the independent variable; but, strictly speaking, anything that occurs in an experimental group after the subjects are randomly assigned can be considered as a part of the independent variable. Some of these common confounds are discussed in sections 2 to 6 below.

2. Testing

Kornhauser administered a test on the first day of class (pretest) and another test at the end of his course (posttest). Attitudes on the posttest

presumably reflected the impact of everything that happened to the student during Kornhauser's course. The posttest should reflect any impact of Kornhauser's attitudes, information transmitted from the textbook, and information transmitted in lecture. But suppose the pretest in itself had some impact on the students' attitudes. Perhaps merely reading and answering the questions on the pretest led some students to reformulate their economic philosophy. Or the test items might have motivated the students to seek out and more closely attend to economic information that might alter their views. It is possible that merely giving the students pretest questionnaires would produce an increase in economic liberalism on a test administered four months later—even if they never enrolled in the economics course! In other words, the pretest itself can be regarded as an aspect or dimension of Kornhauser's independent manipulation.

Effects of pretesting in door-to-door interviewing are even more obvious. What do you think happens after a housewife admits to an interviewer that she does not know the name of her Representative in Congress? Is it likely that she will forget the whole thing at that point? Isn't it more likely that the pollster's question per se motivates her to find out who represents her in Congress? Would it surprise you at all if a second survey of the same sample of housewives found that more of them knew their Representative's name?

At least two solutions are possible. One is to use a posttest-only design. Subjects are randomly assigned to groups, one of which receives the independent manipulation (such as an external threat) and the other does not. Neither group is pretested. However, both groups receive a posttest at the appropriate time. If the posttest scores of the experimental group are different from that of the untreated comparison group, we can safely assume that that difference was caused by the independent manipulation. Since there was no pretest, the independent manipulation would not be confounded with the pretest.

A second solution involves giving a pretest to the control or comparison group. Experimental subjects receive the treatment manipulation plus the pretest, while comparison subjects receive only the pretest. When both groups are posttested at the appropriate time any *difference* between the two groups must result from the independent manipulation. Since both experimental groups received the pretest, the pretest cannot account for a difference between the two groups.[1]

[1] It is possible that the pretest combines or interacts with the treatment to produce certain effects, for example, a certain treatment may only become effective when subject's interest is first piqued by the pretest. A careful experimenter might elect to run two additional conditions in which no pretest was administered. Solomon (1949) and Campbell and Stanley (1963) discuss this four-group design in which the separate effects of the pretest, the treatment variable, and their interactions can be extracted.

In summary, if an independent manipulation is preceded by a testing procedure, we cannot know whether an observed effect results from the independent manipulation alone, the pretest alone, or a combined effect of the two. In order to separate these effects, we must either use a posstest-only design or we must include a comparison group that receives only the pretest and the posttest but does not receive the independent manipulation.

3. History

At the point of random assignment experimental groups should be equivalent within the limits of statistical chance. Anything and everything that happens to the subjects after the point of random assignment can contribute to an observed difference at posttest time. For example, national and international political events having nothing whatsoever to do with Kornhauser's economics course might have increased economic liberalism among University of Chicago students. Thus, "Kornhauser's economics course" was confounded or mixed up with additional historical events, any one of which might have produced the observed effect.

In a true experiment we must carefully examine everything that might happen to the subject from the point of random assignment to the time at which the dependent variable is measured. Any differences in the histories of subjects in various experimental groups can produce an impact on the dependent variable. A solution for this problem is suggested in the following section.

4. Maturation

Maturation refers to growth or changes in behavior that occur simply as the result of passing time. In child development, maturation is thought to be a result of physiological changes that give the child new capabilities. But it is also possible that all college sophomores are generally increasing their economic liberalism with the passage of time. Increased physical size, associated with employment opportunities, might affect his economic attitudes. We can account for Kornhauser's results by assuming that all University of Chicago undergraduate sophomores, as a result of some general maturational process, were becoming more liberal in their economic philosophy.

The solution for maturation and history confounds are similar. Groups of subjects must be randomly divided between an experimental treatment and a comparison or control group. Experimental subjects receive the treatment manipulation or independent variable, and control or comparison subjects do not. At posttest time when the dependent variable is measured, the degree of maturation and changes as a result of history will be equal for experimental and comparison groups. Any differences in economic philosophy between

the experimental and comparison groups must be because of something that happened to the experimental groups but not to the comparison group.

5. Experimenter Expectancy

Most scientists have some sort of commitment to the hypothesis that they are trying to test. Since experimental procedures are often laborious and time consuming, many researchers would not undertake an experiment unless they had confidence that the results would come out as they predict. For example, subjects exposed to high external threat might be observed by an experimenter who hopes or expects that they will become cooperative with each other. The experimenter may willingly or unwillingly communicate his expectations to his subjects. Thus subjects in a high external threat condition might cooperate, not because they were threatened, but because they are living up to the expectations of their experimenter-observer.

Hence, in many experiments there are two differences between experimental conditions: (1) the explicit manipulation (for example, the amount of external threat, and (2) the experimenter's expectations. Since these two variables are confounded, we cannot disentangle their effects.

Robert Rosenthal (1966) has conducted a number of experiments in which he demonstrates that the experimenter can communicate his expectations to subjects in such a way that they fulfill the experimenter-observer's expectations. In one of Rosenthal's tasks, subjects were shown photographs taken from the college yearbook. They were asked to judge how happy the people in the photographs were. Some experimenters were led to believe that the people in the pictures were very happy. Other experimenters were led to believe that the people in the pictures were relatively sad. Even though the experimenters did not say so explicitly, they were able to communicate their expectations to subjects. Subjects run by experimenters who think that the photographs are of happy people rate the pictures as more joyful!

Experimenter expectations create a confound that is both more subtle and more difficult to solve than the previous ones. Ideally, we would have experiments run by experimenters who had no expectations. This might be accomplished by keeping the experimenter "blind" as to what manipulation the subject had received. In the external threat and cooperation study, we might have an external threat manipulation delivered on a tape recording. The experimenter, unaware of what kind of threat manipulation the subjects had received, would then observe the subjects and record their cooperation and competition.

At the very least, it is desirable to have the posttesting or measurement of the dependent variable done by someone who is unaware of the subject's experimental condition. In pencil and paper studies, packets of experimental materials can be passed out to large groups of subjects. The independent

variable is manipulated by placing different instructions in different packets of materials. Since all experimental conditions are run simultaneously and since the experimenter does not know which packet has been given to which subject, experimenter expectations are not confounded with the intended independent manipulation.

6. Demand Characteristics

Martin Orne (1962) has pointed out that many subjects in psychological experiments actively try to discover what the experiment is about. They listen carefully to what the experimenter has to say and closely scrutinize the experimental procedures for cues about the "correct" response. They hope to give the correct response in order to show that they are smart enough to figure the experiment out, or not so stupid as to be "taken in" by a psychological experimenter. Thus a difference between a high and low threat condition might be because, in the external threat conditions, the procedures gave a subtle hint to the subjects that the "correct" response was to cooperate.

Whenever our experimental manipulations give the subject a hint about the "correct" response, demand characteristics are confounded with the experimental manipulation. The clues for the correct response are different in the different experimental conditions; this contrasts with experimenter bias where the confound is in the fact that the *experimenter* behaves differently in the different conditions. With demand characteristics, the confound is in cues contained in the manipulations themselves.

Again the solution for this subtle problem is difficult. About the best that we can do is to closely examine our experimental designs and to try to imagine how it would look from a subject's point of view. Does the experimental manipulation seem to provide an obvious cue as to how the "problem" of the experiment is to be solved "correctly?" If it does, then demand characteristics are confounded with our independent manipulation. Some social psychologists (such as Aronson & Carlsmith, 1968) argue that compelling "cover stories" can be designed to distract the subject and keep him so involved that he does not have time to think about the demand characteristics of the experimental manipulations. Postexperimental questions may also reveal whether the subject responded according to his intuitions about the "correct" response.

GENERALIZABILITY AND EXTERNAL VALIDITY

Up to this point, we have been talking about internal validity (Campbell, 1957; Campbell & Stanley, 1963), which refers to the extent to which we

know what aspect of the independent manipulation produced the observed difference. The confounds discussed in the previous section are all threats to internal validity.

But problems of external validity or generality also plague the experimental social psychologist. Perhaps Kornhauser's results did reflect the different information transmitted in text and lecture *for University of Chicago undergraduates in the late 1920s.* But would the same textbook have produced attitude change at Northwestern University 30 miles to the north? Would the same information have produced attitude change among members of the local chamber of commerce? Or among members of the local carpenter's union? These questions about the *generalizability* of a study all refer to *external validity.* Given that we understand adequately how the results were produced in this particular setting with this particular group of subjects, to what other settings and what other groups of subjects are the results applicable? Also, would we get the same results if we tested the same hypothesis with another method? Similar questions can be asked of all empirical studies. Do they hold only for that particular experimenter, the particular method of operationalizing the independent and dependent variables, the cover story, subject population, the time of day, week, year, and so on?

Generalizability and external validity are increased by replications. Even efforts at "exact replication" must change some aspects of the procedure; different subjects participate, current events change. Generalizability is maximally increased when the replication shares only the theoretical if-then hypothesis with the original. If two experiments differ in as many respects as possible (such as subject population, attitude issue, experimenter, and operational definitions of variables) and share little more than the theoretical hypothesis, then we have demonstrated that the results, originally demonstrated in one context or situation, can be generalized to another context or situation.

We can mention one other major problem of external validity. Most subjects in social-psychological experiments know that they are "in an experiment." Ethical considerations make us think twice about applying an independent manipulation to subjects who are not aware of the fact that they are in an experiment. But would the results generalize to other settings where the individuals do not think that they are "participating in an experiment?"

This may be the most difficult of all the problems raised so far. One solution is to conduct experiments in settings that the subject has not explicitly labeled as "experiments." Since traditional questionnaire and interview methods sensitize subjects to the fact they are being studied, we need to develop "unobtrusive" measures that can gather data about the

subject without interfering with or biasing his normal psychological processes (see Exercise 3). For example, the popularity of a museum exhibit might be measured by the rapidity with which the floor tiles are worn and need replacing surrounding the exhibit instead of by giving a questionnaire to each passerby. Or the popularity of a television program might be observed by noting the drop in water pressure during commercial breaks (when viewers flush the toilet and fill their glasses) (Webb, Campbell, Sechrest, & Schwartz, 1966).

RELIABLE MEASURES OF THE DEPENDENT VARIABLE(S)

If an instructor gives 30 students a yardstick and asks them to measure his desk to the nearest 1/32 of an inch, he may get 30 different answers. Numbers yielded by our measuring instruments contain a component of random error. The less the random error (the greater the *reliability* of the dependent measure), the greater is the probability that the measuring instrument can detect a difference produced by the independent manipulation. Exercise 2 on attitude measurement demonstrates a technique to increase the reliability of the dependent variable.

Psychological measurement and testing are complex topics and they are often the subject of entire undergraduate courses; so you will have to take the following recommendations on faith. The reliability of a dependent variable generally increases with: (1) A greater number of separate measurements. The subjects can be measured by a large number of items, by a large number of observers, or with a large number of separate measurement techniques. (2) A high correlation among the separate measurement techniques. (3) A minimum methodological overlap among the various measurement procedures.

CONCLUSION

No chapter "talking about" research is a substitute for actually doing research. This introductory chapter was written to introduce and illustrate some of the problems that often stand between the social psychologist's questions and his answers.

The chapter began with a discussion of the five steps in scientific problem-solving: (1) Select a problem area, situation, or phenomenon (2) formulate a specific if-then hypothesis, (3) construct a literary definition for both variables in the if-then hypothesis, (4) choose a method to quantify and measure the two variables in the if-then hypothesis, (5) select a research or question-answering strategy. The next section dealt with random error and

statistics. The last major section dealt with confounds in experimental designs which make unequivocal interpretation of the results impossible. The last two sections introduced the problem of generality and stressed the importance of accuracy and reliability in measuring the dependent variable.

References

Aronson, E., and Carlsmith, J. M. Experimentation in social psychology. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology, Volume II*. Reading, Mass.: Addison-Wesley, 1968.

Campbell, D. T. Factors relevant to the validity of experiments in social settings. *Psychological Bulletin,* 1957, **54**, 297–312.

Campbell, D. T., and Stanley, J. C. Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), *Handbook of research on teaching*. Chicago, Ill.: Rand McNally, 1963. Pp. 171–246. Also available in paperback: Campbell, D. T., and Stanley, J. C. *Experimental and quasi-experimental design for research.* Chicago, Ill.: Rand McNally, 1963.

Kornhauser, A. W., Changes in the information and attitudes of students in an economics course. *Journal of Educational Research,* 1930, **22**, 288–298.

Latané, B., and Darley, J. M. *The unresponsive bystander: Why doesn't he help?* New York: Appleton-Century-Crofts, 1970.

Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist,* 1962, **17**, 776–783.

Rosenthal, R. *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts, 1966.

Solomon, R. An extension of control group design. *Psychological Bulletin*, 1949, **46**, 137–150.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L. *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago, Ill.: Rand McNally, 1966.

Exercise

Interviewing

This exercise is designed to demonstrate some important methodological problems that may arise in naturalistic interviewing. The experiment requires two persons, so choose a partner, and flip a coin to determine which role, "Supervisor" or "Interviewer," each of you will assume. Because this demonstration will not be useful if you are informed about the entire procedure in advance, *separate* instructions are provided for the supervisor and the interviewer. Each partner *should read only his own instructions*; do not read those of your partner.

With your partner, select a controversial current issue. An example of an appropriate topic is a federal-government involvement in the affairs of others (for example, foreign aid, welfare programs, support of educational institutions, disseminating birth control information). Devise a set of five questions for your topic. Try to word about half of the questions so that agreement indicates one side of the attitude dimension; phrase the rest of the questions so that agreement indicates the opposite side. For the example, agreement with questions 1 and 3 signifies a pro-welfare attitude, but agreement with questions 2, 4, and 5 signifies an anti-welfare attitude. In addition, include one question about the respondents' political affiliations. An example of an interviewing schedule to determine the respondents' attitudes towards federal welfare programs is presented below.

Example Interview Schedule

ATTITUDES TOWARDS FEDERAL WELFARE PROGRAMS

1. Do you believe that the federal government (Washington) should increase the money given to the underprivileged?
   A. Very much agree that they should increase it.
   B. Somewhat agree they should increase it.
   C. Somewhat disagree they should increase it.
   D. Strongly disagree they should increase it.

2. Much federal financial aid to the unemployed tends to rob these people of individual initiative to become self-reliant.
   A. Agree very much with this statement.
   B. Agree somewhat with this position.
   C. Disagree somewhat with this position.
   D. Disagree very much. Doesn't rob them of any initiative at all.

3. Supporting the unemployed does not produce excessive burdens on the taxpayer.
   A. Agree very much.
   B. Agree somewhat.
   C. Disagree somewhat.
   D. Disagree very much.

4. The federal government's support of the poor is perhaps the first step in producing a too powerful federal government.
   A. Agree very much.
   B. Agree somewhat.
   C. Disagree somewhat.
   D. Disagree very much.

5. The support and help to the poor should be a matter for the local government (city or state); it is not Washington's business.
   A. Agree very much.
   B. Agree somewhat.
   C. Disagree somewhat.
   D. Disagree very much.

6. Do you typically vote (or think you will vote) for a Republican or a Democratic candidate in national elections?
   A. Democratic
   B. Republican
   C. Independent
   D. Don't know, refuse to answer.

Procedure

With your partner, find a field setting that will enable you to interview a number of *males*, varying in age and social class. Most downtown street corners will be suitable for this purpose. After selecting the setting, begin interviewing males on the selected topic. Now turn to your own special instructions and data forms. Do not read the data analysis, discussion, or the other person's instructions until the data are collected.

Instructions to Interviewer

You are to choose 25 male pedestrians who will serve as respondents in your survey. These males must be alone, that is, not be members of a group. The reason for this restriction is that if you attempt to interview males within a group context, their answers might be influenced directly or subtly by the presence of other members of the group. Since many interviewers get better results if they interview only those subjects with whom they feel compatible, you are to select 25 respondents with *whom you might feel comfortable*. After completing the survey, you will receive further instructions from your "supervisor" partner.

A word of caution: some of your respondents may give lengthy discourses and otherwise evade reducing their answers to the choices you give them. Encourage them to summarize their positions by indicating which of the alternatives come closest to their general attitudes. You might say, "I'm only allowed to mark one of these choices," and then repeat the alternatives. A data sheet for recording the interviewee responses is provided on page 37. DO NOT READ FURTHER UNTIL YOU HAVE COMPLETED THE EXERCISE.

Instructions for Supervisor

Do not inform your partner about these instructions.

The purpose of the experiment will be to determine the effects of interviewer selection procedures on the results obtained. As you may have read elsewhere, adequate sampling requires that selection should not be affected by the whims and fancies of the interviewer. In a valid random sample, each potential interviewee has an equal chance of being selected. It has been found that interviewers, if allowed, will tend to pick respondents who may yield very different answers on many important issues than would a sample of randomly selected persons. This exercise is designed to demonstrate this phenomenon. In fact the interviewer has been instructed to select subjects with whom he anticipates feeling most comfortable. (You may read the Instructions to Interviewer if you wish.)

When you arrive at the field setting decided on by you and your partner, let your partner decide in any manner he wishes who will be the first 25 male respondents. You will record data about the characteristics of the respondents he selects. He has been told to expect further instructions from you after he completes his first 25 interviews. After he has completed these 25 interviews, you will select a second sample of 25 respondents as follows.

Second Sample Selection

The respondents are to be unaccompanied males. Do not select respondents from a group of pedestrians; only select males who are alone. The method you use for selection, as compared to your partner's, defines the two treatments of this experiment. You will choose some *arbitrary* rule for selecting your sample. For example, you may assign your partner to interview each third, fifth, or tenth eligible male that approaches after termination of the previous interview. Or you may assign your partner to interview the first eligible male that passes after a 30-second period has elapsed from the termination of the previous interview. There are two necessary requirements for your sampling procedure. First, selection must be independent of your individual preferences. All unaccompanied males passing the interview site should have an *equal chance* of being selected. Do not eliminate members of minority groups, the poorly dressed, handicapped, etc. Second, your standards of selection should remain constant throughout the sampling. No potential respondents should be excluded from the survey except on the basis of the rules you have decided on before starting the sampling procedure.

Data Recording

In addition to assigning the second 25 interviewees, you will record data concerning their characteristics. Two separate data forms are provided; one for the sample the interviewer selects, the other for the sample you select. Some characteristics on which your sample and your partner's may differ might be the respondents' race, social class (as estimated by dress) and age. You may add other characteristics, but be sure that they are easily identified so that your ratings on that characteristic may be assumed to be highly reliable; that is, you can be confident that others would agree with your classifications. After the second 25 interviews have been completed, ask your partner to read your supervisor's instructions. Then both of you may proceed to the following Data Analysis section.

Neither Interviewer or Supervisor should read the following until interviewing is completed.

Data Analysis

The experiment was conducted to assess the effects of differences in sampling procedures. In the initial sampling procedures, the interviewer was allowed to sample those subjects with whom he anticipated being most comfortable. If the interviewer has such prerogatives in the sampling of respondents, he often tends to choose individuals who fit the image of white Protestant middle-class Americans. (If the interviewer himself is a white Protestant middle-class American.) Has this been the case with the interviewer? Did he tend to select members of the upper social class or Caucasians in greater numbers than yielded by the supervisor's arbitrary selection procedure? Tally the characteristics of both samples. What differences are there? To assess whether the two samples vary according to social class, you can compare them by means of a chi-square analysis (Appendix B). The data may look like this, for example.

	Type of Respondent	
	White Collar	Blue Collar
Interviewer-selected sample	20	30
Supervisor-selected sample	30	20

chi-square = 3.84 $p < .05$

A similar analysis might be computed for differences in race. Race bias in selecting the samples can be analyzed by chi-square as follows.

	Type of Respondent	
	White	Black
Interviewer-selected sample	40	10
Supervisor-selected sample	30	20

Chi-square = 4.85 $p < .01$. (Notations such as $p < .05$ and $p < .01$ are explained in the Introduction, pp. 15-16.)

Effects of Sampling on Interview Data

The next question is whether the different sampling techniques produced differences in attitudes of the two samples. Sample characteristics could be quite different, yet attitudes could be similar; or the two samples might appear similar and yet yield different attitude patterns. First, score each respondent's answers by assigning a score of 4 to strongly favorable, 3 to mildly favorable, 2 for mildly against, and 1 to strongly against.

In order to prevent response bias, it may be wise to word some of your attitude items so that a "strongly favorable" response indicates opposition to the topic of interest. In the Example Interview Schedule, a favorable response to item 2 indicates opposition to welfare programs. Obviously, such items must be reverse-scored, for example, strongly favorable would be assigned a score of 1. Each respondent should then be assigned a total score—by adding scores for all of his responses except for the last question concerning party identification. Calculate the average or mean attitude within each sample by adding each individual's total score and dividing by the number of respondents or 25. What is the average attitude toward this issue for these two samples? Do attitudes for the two samples differ? You can test the significance of this difference by means of a *t*-test (see Appendix A).

The final question, which concerns party affiliation, can be used as another measure of interviewee differences produced by variation in sampling procedure. Responses to this question can be analyzed, like those of race and social class judgments, by means of a chi-square analysis (see Appendix B). If there are such sampling differences, what might have been the cues that affected selection by the interviewer? Did you notice differences in hair style, dress, race, or social class that might be related to the respondents' responses to this item and which may have cued the interviewer in his selection of such subjects?

If there is a difference in the attitudes of the two samples, is it predictable from the differences, if any, found in the personal characteristics of the samples (for example, race, social class, political affiliation)? In other words, are the personal characteristics that distinguish your two samples correlated with the attitudes that were measured?

Discussion

This exercise was devised to indicate one of the problems associated with adequate sampling, that is, the personal biases of the interviewer. Given that you found differences in the samples selected, it is possible that interviewer bias may explain such a finding. Can you think of other explanations? One possibility is that the samples were drawn at different times of the day and that the kind of pedestrian interviewed in the two samples may differ in some critical respects. Is this a possibility in your case? What other plausible rival hypothesis can you think of that might account for such differences?

What are the limits of your attitude data in generalizing to the general public? What might be some characteristics of unaccompanied male pedestrians that may affect their answers in a way that would limit generalization to the general population?

References

Hyman, H. H., Cobb, W. J., Feldman, J. J., Hart, C. W., and Stember, C. H. *Interviewing in social research.* Chicago: University of Chicago Press, 1954.

This book reports a number of investigations concerned with sources of interviewing error. Such topics as situational effects on responses, interviewer characteristics and their impact on the respondent, and sources and control of errors are discussed. A most important review for those interested in problems associated with interviewing techniques.

Rosenthal, R. *Experimenter effects in behavioral research.* New York: Appleton-Century-Crofts, 1966.

This book summarizes and integrates the vast number of studies concerning the impact of the experimenter on the conduct of subjects within laboratory experiments.

"Science and politics: AMA attacked for use of disputed survey in 'Medicare' lobbying." *Science*, 1960, **132**, 604–605.

This article is a debate about the sampling techniques employed in a survey concerning medical coverage. The study was employed by the American Medical Association and probably had important political implications. For those interested in uses and abuses associated with survey samples, this is a relevant article.

Interviewee Responses Attitude Survey
First Sample
(to be filled out by interviewer)

Code responses such that favorable statements receive the highest scores and unfavorable statements receive the lowest scores. In the example on federal welfare programs, selecting the "A" alternative in question 1 would be a scored 1, B-2, C-3 and D-4.

Respondent No.	Items: 1	2	3	4	5	Total	Party Affiliation: Republican	Democrat	Other
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
TOTAL									
MEAN (DIVIDE TOTAL BY 25)									

Interviewee Responses Attitude Survey
Second Sample
(to be filled out by supervisor)

Code responses such that favorable statements receive the highest scores and unfavorable statements receive the lowest scores. In the example on federal welfare programs, selecting the "A" alternative in question 1 would be a scored 1, B-2, C-3 and D-4.

	Items						Party Affiliation		
Respondent No.	1	2	3	4	5	Total	Republican	Democrat	Other
1									
2									
3									
4									
5									
6									
7									
8									
9									
10									
11									
12									
13									
14									
15									
16									
17									
18									
19									
20									
21									
22									
23									
24									
25									
TOTAL									
MEAN (DIVIDE TOTAL BY 25)									

Interviewer
Data Sheet—Attitude Survey
(to be filled out by supervisor)

Date of Survey ____________ Time: From ___ To _____ Location ______________

	Race			Dress			Estimated Age					Party Affiliation		
Respondent No.	Wh	Bl	Other	Wearing tie (White Collar)	Work clothes (Blue Collar)	Other	15-24	25-34	35-49	50-64	65 +	Rep.	Dem.	Other
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
12														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
TOTAL														

Supervisor
Data Sheet—Attitude Survey
(to be filled out by supervisor)

Date of Survey ______________ Time: From ____ To _____ Location ____________________

	Race			Dress			Estimated Age					Party Affiliation		
Respondent No.	Wh	Bl	Other	Wearing tie (White Collar)	Work clothes (Blue Collar)	Other	15-24	25-34	35-49	50-64	65 +	Rep.	Dem.	Other
1														
2														
3														
4														
5														
6														
7														
8														
9														
10														
11														
13														
13														
14														
15														
16														
17														
18														
19														
20														
21														
22														
23														
24														
25														
TOTAL														

Exercise 2

Measuring Attitudes

This exercise is designed to give you practice in constructing a scale to measure attitudes towards a specific topic. First you will select statements for inclusion in the scale (construction phase); then you will test the scale to determine whether it discriminates between groups that might be expected to hold different opinions[1] (validation phase). Both construction and validation techniques are adapted from the classic scaling method introduced by L. L. Thurstone (1928, 1929).

In the construction phase you will ask a number of people to judge the items you propose to include in your scale. These "judges" will estimate how positive or negative each item is toward the attitude topic. The data gathered from the judges will allow you to select items that represent different points along a scale of attitudes ranging from very favorable to very unfavorable. Later in the validation phase you will evaluate your subjects' opinions by placing them on the scale that was constructed according to the data collected from the judges. In order that subject attitudes are placed in the appropriate context, both subjects and judges must be drawn from the same population.

Construction

Choose an issue or topic that interests you. Controversial current issues such as legalizing marijuana or abortion, and the desirability of college

[1] Although the term "opinion" is often used to mean overt expression of an underlying "attitude," the two words will be used more or less interchangeably in this exercise. We shall also ignore here discrepancies between self-reported attitudes and related behavior. This topic is discussed in Exercise 3 on Unobtrusive Measures.

fraternities are especially good for this purpose. Generate 20 brief statements that seem to range from very favorable to very unfavorable on your topic. Some examples of statements about college fraternities follow.[2]

1. It is so hard to decide whether fraternities are good or not.

2. Fraternities are fine for certain types of people.

3. All in all, there is more wrong than right about the fraternity system.

4. The sooner college fraternities cease to exist the better.

5. The nonfraternity student is missing one of the most important aspects of college training.

6. There is so much to find fault with in the institution of college fraternities.

7. Fraternities do far more harm than good.

8. Much can be said both in favor and against college fraternities.

9. Fraternity membership usually results in severe limitation on originality and productivity.

10. A large part of the opposition to fraternities is the result of jealousy and ignorance on the part of those who failed to gain admission to a fraternity.

11. College fraternities provide healthful experiences for many people.

12. It would be a calamity if fraternities didn't exist.

13. The question of whether college fraternities are a good thing or not is petty and unimportant.

14. Fraternities are loafer's clubs.

15. It is difficult to understand how anyone can take an unqualified stand on the question of fraternities.

16. Fraternities are absolutely terrible.

17. Fraternities help to mold the college man into a well adjusted and productive member of society.

18. An employer would be wise to give preference to a fraternity man.

19. College fraternities are among the finest institutions yet devised by Western man.

20. Fraternities may be all right, but a lot of changes are necessary before they can be given wholehearted approval.

In choosing statements about your topic avoid vague wording or compound sentences that include two separate thoughts. Try to distribute your

[2] These statements used with permission from T. J. Banta, Social attitude and response styles. *Educational and Psychological Measurement,* 1969, **21**, 543-557.

statements so that about half are worded positively toward your topic and half negatively. If you cannot invent enough statements, seek help from friends and books. Write the statements on index cards, then ask at least five judges to sort the statements into seven stacks. It is easy to recruit judges for this purpose at restaurants and lounges frequented by college students. Prepare seven cards to identify the stacks and place them in a row on a table as illustrated. You may want to tape the cards in place.

1	2	3	4	5	6	7
Strongly in favor	Moderately in favor	Slightly in favor	Neutral	Slightly Opposed	Moderately Opposed	Strongly Opposed

Instruct your judges as follows:

"Sort the statements into seven stacks. Put those statements which are most favorable to (topic) in stack 1 and those most unfavorable to (topic) in stack 7. Less favorable and unfavorable statements should be placed in the intermediate stacks. Try to place approximately the same number of statements in each stack. After you sort the statements the first time, check through each stack and make any changes you wish. Remember YOU ARE *NOT* SORTING ACCORDING TO YOUR PERSONAL OPINION OF THE TOPIC. YOU ARE TO SORT ACCORDING TO WHAT KIND OF ATTITUDE THE STATEMENT REPRESENTS. If the topic was baseball, the statement 'Baseball is the worst game ever invented' would usually be sorted in stack 7 indicating strong opposition regardless of your own feelings about baseball. Even an ardent baseball fan would judge that kind of statement as *unfavorable* to baseball. Any questions? (Pause) OK, go ahead."

You can adapt these general instructions to your own style. Be sure that the judge checks through the stacks after sorting to see if he wants to make any changes.

Each of your statements should be identified by a code number written on the back of the card. Use the attached data sheets to indicate the stack in which each statement was sorted. After all the judgments are collected, sum the column of X values (stack number into which statement was sorted) for each statement and divide this total by the number of judges (N) to obtain a mean judgment score (M). $M = X/N$.

One criterion for including statements in the scale is that judges should generally agree on the rating of the statement. To calculate judge agreement obtain deviation scores (d) and calculate the average deviation (AD) for each statement. Work space is provided on the data sheet. First calculate d for each item by subtracting X from M. Disregard minus signs ($d = |M - X|$). Average deviation is simply the mean of the d's for each item: $AD = d/N$

Calculation Example

	Statement Number							
	1		2		3		4	
Judge	X	d	X	d	X	d	X	d
A	7	0	4	0.3	1	2.7	7	0.3
B	7	0	3	0.7	4	0.3	6	0.7
C	7	0	4	0.3	6	2.3	7	0.3
Total (ΣX)	21		11		11		20	
Mean $\left(\frac{\Sigma X}{N}\right)$	7		3.7		3.7		6.7	
Total Deviation (Σd)	0		1.3		5.3		1.3	
Average Deviation $\left(\frac{\Sigma d}{N}\right)$	0		.43		1.8		.43	
√ = included in scale	√		√					

X = score stack number judges assign to each statement

N = number of judges

ΣX = total of scores assigned to a statement by all judges

Mean $\left(\frac{\Sigma X}{N}\right)$ = average score assigned to a statement

Deviation (d) = $/M - X/$ (treat all deviations as positive values, i.e., disregard minus signs)

d = total of deviations

where N = number of judges. Statements with maximum agreement or minimum variation between judges will have the lowest *AD* scores.

Select approximately 7 or 8 statements for your attitude scale. Ideally you will be able to find statements with: (1) high judge agreement or low *AD* scores and (2) mean values fairly evenly distributed across the possible range. You may be able to construct a questionnaïre with one item representing each of the 7 scale positions.

How to calculate mean judgment scores and average deviations is illustrated in the Calculation Example. In this simplified example there are only three judges sorting four statements which are numbered from 1 to 4.

Judge A sorts statement 1 into the seventh stack indicating that he thinks the statement is strongly unfavorable to the topic. He sorts statements 2, 3,

and 4 into stacks 4, 1, and 7 respectively. Similar X values corresponding to ratings by Judges B and C are entered in the second and third rows. After all the ratings have been entered, the columns representing each of the four statements are totalled (ΣX = sum of the X values). The total for the first statement is 21 since all judges rated it 7—strongly unfavorable. Next, the mean judgment score is calculated by dividing ΣX by the number of judges, in this case 3. After the mean score is known, deviation scores (d) can be entered. The difference between each X value and the mean judgment score is d. For statement 3, judge A's rating is below the mean, but judges B and C rate above the mean. Notice that deviation scores are always positive whether they represent differences above or below the mean. In the example, the deviations are added and divided by the number of judges (3) to compute the average deviation. Statements 1 and 2 are checked for inclusion in the scale, since the agreement between judges is good (perfect for statement 1) and since they represent different points on the scale. Statement 3 is omitted because it has a high average deviation indicating low judge agreement; statement 4 is not included because its mean judgment score is very close to the score for statement 1.

Validation

An attitude scale can be validated by administering it to two or more large samples that are assumed to be different from each other on the relevant dimension that the scale presumes to measure. There are many criteria for selecting validating samples; however, choice of sample should be guided by the purpose of the scale. For example, a scale on attitudes toward college fraternities could be administered to fraternity men and dormitory men; a scale on attitudes towards marijuana could be tested on marijuana users and nonusers. A test designed to diagnose mental illness might be validated by comparing scores obtained by hospitalized schizophrenics with scores obtained from people who had never been hospitalized for mental illness.

It will not be practical for you to obtain two large samples, so for this exercise you may test your scale on only 20 subjects. Select 10 who are in favor of the position being assessed and 10 more who are opposed. You might test pro-military versus anti-military persons, an atheist and a church activist on a religiosity scale, or a hunter versus a nonhunter on attitudes toward firearm legislation.

Present the subjects with the 7 or 8 statements you have selected for your scale. Ask them to indicate their own personal agreement or disagreement with each statement; that is, the subject should sort the statements into two piles: (1) those that express his own sentiments or attitudes towards the topic and (2) those that deviate from his own position in either direction.

Calculate the attitude score for each subject. First list the scale value for

each statement agreed with for each subject. Use Table 1 for recording data. Arrange the values in order from lowest to highest, or most favorable to least favorable. The scale values are the mean judgment scores (*M*) that you calculated during the construction phase. If the subject agrees with an odd number of statements, his attitude score is simply the value of the middle (median) statement he agrees with. If your subject agrees with an even number of statements, his score is the midpoint of the scale distance between the two middle statements. For example, if your subject agrees with four statements whose values are 1.3, 5.0, 6.0, and 6.2, his attitude score would be 5.5 or the midpoint between the two middle values. This median method of computing attitude score does not reflect extreme values as much as would a method based on arithmetic mean.

A *t*-test (Appendix A) can be applied to determine if the attitude scores for two validating groups were different from each other beyond what might reasonably be expected by chance.

Discussion

How did you select the two groups of subjects? Did they differ as you expected they would? Could your scale be validated with other groups? A common student reaction to this validation technique is "So what? Obviously one group will respond to scale items differently from another group with opposing interests." In many cases, especially when validating groups are selected according to a clear criterion, e.g., clergymen versus atheists, it may indeed seem obvious when differing responses are obtained. It should be remembered, however, that once the scale is validated by testing people with known attitudes, people with unknown attitudes or moderate attitudes can be tested and compared using the same measuring instrument.

Such scales are also useful for testing the same person at different times. This procedure is followed by many attitude change experimenters who typically test their subjects' attitudes, then expose them to a persuasive communication, and finally retest the subjects on the same scale to determine amount of change. An improvement on this technique is to use one scale to measure initial attitudes and a different but equivalent scale after the persuasion attempt. If the second scale is equated with the first for mean judgment scores, change can be measured while avoiding problems associated with using the same scale twice within a short time. One such problem is that subjects will remember their initial responses and then attempt to appear consistent even when their attitudes have been affected by the experimental treatment.

Although Thurstone scaling has proven useful in studying attitudes, the method and its assumptions have been frequently criticized on several grounds. First it has been found that the construction of the scale can be

affected by the attitudes of the experimenter who makes up the initial pool of the items, and by the attitudes of the judges who sort the items in the construction stage. The experimenter's own attitude can bias his selection of items for the judges to sort. This experimenter bias can best be eliminated by selecting a very large number of items from many sources. A more serious problem is that the judges' own attitudes can bias the way in which cards are sorted, consequently affecting mean judgment values. Thurstone himself (1928) has stated that for a scale to be valid "the scale values of the statements should not be affected by the opinions of the people who help to construct it." In fact it has been found that judge opinion does affect these values, for example, an ardent baseball fan might indeed sort statements differently than an antisport individual despite specific instructions to ignore personal opinion. Judge bias cannot be eliminated in many cases; however, it is usually possible to obtain a representative sample of judges whose opinions are similar in intensity and range to those of the subject population.

If such a representative group of judges (who cannot sort independently of their own opinion) is obtained, the mean judgment values for each scale item will be similar to values obtained from ideal unbiased judges; however, because biased judges will tend to sort the same statement in different stacks, agreement between judges will be lower yielding a higher average deviation. Thus judge bias will not affect the average or mean judgment value if the judges are representative, but judge agreement on particular items will be affected. As an additional exercise you might select two distinctly different samples of judges, ask them to sort attitude statements using the instructional format on p. 47, and then determine if judge attitude is related to the supposedly objective sorting behavior, that is, does one group of judges rate the same items differently than another group of judges?

References

Thurstone, L. L. Attitudes can be measured. *American Journal of Sociology*, 1928, **33**, 529–544. Also reprinted in Fishbein, M. (Ed.) *Readings in attitude theory and measurement,* New York: John Wiley, 1967, 77–89.

In this article, Thurstone describes the theoretical assumptions underlying his method. He also discusses validation methods and criteria for determining ambiguous or irrelevant items.

Thurstone, L. L., and Chave, E. J. *The measurement of attitude.* Chicago: University of Chicago Press, 1929 (as described by Edwards, A. L., *Techniques of attitude scale construction*), New York: Appleton-Century-Crofts, 1957, 83–119.

The Thurstone-Chave book describes the Thurstone scale construction technique in detail as it is applied to attitudes toward the church. Edwards lucidly describes the scaling technique and some of its major criticisms.

Data Sheet 1—Statement Numbers 1-10

	1		2		3		4		5		6		7		8		9		10	
Judge	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d
1																				
2																				
3																				
4																				
5																				
6																				
7																				
Total (ΣX)																				
Mean $\left(\frac{\Sigma X}{N}\right)$																				
Total Deviation (Σd)																				
Avg. Dev. $\frac{\Sigma d}{N}$																				
= included in scale																				

X = score stack number judges assign to each statement

N = number of judges

ΣX = total of scores assigned to a statement by all judges

Mean $\left(\frac{\Sigma X}{N}\right)$ = average score assigned to a statement

Deviation (d) = $|M - X|$ – (treat all deviations as positive values, i.e., disregard minus signs)

Σd = total of deviations

Data Sheet 2—Statement Numbers 11-20

	11		12		13		14		15		16		17		18		19		20	
Judge	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d	X	d
1																				
2																				
2																				
4																				
5																				
6																				
7																				
Total (ΣX)																				
Mean $\left(\frac{\Sigma X}{N}\right)$																				
Total Deviation (Σd)																				
Avg. Dev. $\frac{\Sigma d}{N}$																				
= included in scale																				

X = score stack number judges assign to each statement

N = number of judges

ΣX = total of scores assigned to a statement by all judges

Mean $\left(\frac{\Sigma X}{N}\right)$ = average score assigned to a statement

Deviation (d) = $|M - X|$ – (treat all deviations as positive values, i.e., disregard minus signs)

Σd = total of deviations

Data Sheet 3

Subject	Group I Items Agreed with										Score	Group II Items Agreed with										Score
1																						
2																						
3																						
4																						
5																						
6																						
7																						
8																						
9																						
10																						
11																						
12																						
13																						
14																						
15																						
16																						
17																						
18																						
19																						
20																						

Mean Score For Group I: ____

Mean Score For Group II: ____

Exercise 3

Unobtrusive Measures

This exercise is designed to illustrate nonreactive measures and to demonstrate how such measures relate to more traditional questionnaire and interview techniques. Nonreactive or unobtrusive measures (Webb, Campbell, Schwartz, and Sechrest, 1967) attempt to measure the reactions of subjects who are unaware that they are being studied. Subjects who are aware that they are being observed or studied may react self-consciously, and self-conscious subjects may behave atypically or they may distort their actual beliefs. For example, a racist subject who feels that his true beliefs would not be socially acceptable or proper in the context of a research interview may be hesitant to express his real feelings to a researcher.

In the past psychologists have relied heavily on projective testing to detect areas in which a subject may be sensitive about revealing his attitudes. Projective testing is based on the assumption that the subject will project his real attitudes into a test situation that requires him to invent a story about a stimulus such as a picture. However, because the interpretation of these tests has been questioned and because administration and scoring are time-consuming, they are losing favor with attitude researchers. Moreover, projective tests may also carry reactive components of their own.

In addition to reducing measurement bias, nonreactive measures can increase a researcher's confidence in his inferences based upon other methods of assessment. Each assessment technique may have peculiar features which reduce the level of confidence that can be placed in the results. For example, some people tend to respond with true (or false) answers to true-false items on an attitude questionnaire regardless of the question that is asked. A score on such a true-false survey is determined only in part by the

individual's real attitudes; his score is also affected by his tendency to answer most questions true (or false) whether the answers truly reflect his attitudes. One means of improving confidence in attitude measurement is to "triangulate" methods, that is, to employ several different techniques to measure the same attitude. If your neighbor denies that he holds racist attitudes, you might gain additional confidence about his supposed brotherly feelings if you observe him actually interacting with someone of another race. Such nonreactive behavioral observations, when used in conjunction with more traditional verbal report techniques, add considerable validity to inferences concerning another's attitude.

Although unobtrusive measures offer distinct advantages, especially in reducing unwanted effects of subject awareness, it is not always possible or desirable to utilize these techniques. Sometimes unobtrusive data collection requires long tedious hours of observation or expensive recording equipment, and even more importantly, studying a respondent without his express consent can involve serious ethical considerations. To what extent is your invasion of an unknowing subject's privacy justified by the value of the data? How are the subject's rights and welfare safeguarded by your procedures? These questions must be answered before unobtrusive measures are employed.

Procedure

Your task will be to develop and validate an unobtrusive measure. Select two groups of people who probably differ considerably on some interesting dimension and who also can be differentiated on the basis of an unobtrusive measure. For example, style of dress might be related to academic major or occupational interests; certainly a young man with a slide rule attached to his belt is more likely to be an engineer than an artist. Webb et al. (1967) have suggested a more subtle relationship between appearances and interests—that hard psychologists (those interested in experimental or physiological) wear their hair in shorter styles than those interested in clinical or social psychology. Hair styles might be more generally related to attitudes, for example, it is often assumed that bearded or long-haired college students differ from their clean-shaven crew-cut colleagues on several attitude dimensions. The examples of hair style or dress are merely suggestions of possible unobtrusive measures. You are free to select your own measure and a means of validating that measure.

Consider some personal characteristic that might differentiate students with varying attitudes, interests, or living habits. Do you think that beards and long hair are more frequently displayed by students with interests in the humanities and the social sciences than by those involved in the hard sciences such as chemistry and physics? Perhaps a particular kind of hair

style might identify fraternity men. When you have chosen an easily visible personal characteristic which may differentiate among students with varying interests, select a research site that will allow you to observe people who can reasonably be assumed to be members of a group you are studying. For example, if you hypothesize that long-haired college males are more likely to have humanitarian interests than their short-haired peers, first station yourself outside a social work class, then outside an engineering class. Or if you decide to compare fraternity to dormitory residents, locate yourself at the entrance to their respective living units.

If you are measuring a personal characteristic that is somewhat difficult to define (for example, just what is long versus short hair?), you will need to conduct a reliability study to determine whether two or more independent observers agree on what is meant by the concept. In order to establish reliability, at least two observers who are unaware of each other's ratings must demonstrate substantial agreement in their judgments of the subjects of the dimension of concern. Ask two friends or classmates to rate independently 20 or more subjects on the characteristics you plan to use in your study. If they agree at least 75% of the time you may proceed with your data collection. If they do not agree in the great majority of instances, why don't they? Reliability can be improved by explicitly spelling out the criteria that are used to make the measurement. How long is long hair? Just what kinds of clothes indicate that a student should be classified as well-groomed? If you believe that better training or a better rating system will improve reliability, retrain your observers and have them discuss their previous ratings before allowing them to make additional ratings. If they improve to 75% agreement, or any other criterion that you have decided is appropriate for your study, you may start collecting data. See exercise 11 on group processes for a further discussion of reliability. If the characteristics that you want to unobtrusively assess can be easily defined and observed (such as engagement rings and slide rules), no reliability study will be necessary.

After reliability has been established, select 50 subjects from each of the two types of people who are predicted to exhibit differences on the characteristics assessed and begin your observations. One of the raters involved in the reliability test must be used in making the observations.

Data Analysis

Record the observations on the Data Sheet, p. 67. You can determine whether your measure differentiated between the two groups in the manner you predicted by performing a chi-square analysis (Appendix B) on the total frequencies entered in the table at the bottom of the Data Sheet.

References

Campbell, D. T., and Fiske, D. W. Convergent and discriminant validation by multitrait-multimethod matrix. *Psychological Bulletin*, 1959, **56**, 81–105.

These authors theoretically describe the relationships between several different traits or behaviors, each of which is measured by several different methods. An examination of these relationships allows the researchers to know how much of a specific score actually reflects the trait that is being measured and how much of the score is a result of only the testing or measuring method itself.

Messick, S. and Jackson, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychological Bulletin*, 1961, **58**, 299–304.

This report reviews early investigations on the role of response acquiescence on true-false answers to a particular test. The article launched a series of experiments concerning the propensity of some individuals to answer true or false to items irrespective of the content of the test items.

Webb, E. J., Campbell, D. T., Schwartz, R. D., and Sechrest, L., *Unobtrusive measures: Nonreactive research in the social sciences.* Chicago: Rand McNally & Co., 1967.

This book presents arguments for the greater use of unobtrusive measures and reviews those that have been employed in the past and those that may be employed in the future. Undergraduate students will find this book enjoyable and easy to read.

Data Sheet

Subject Sample Type 1 Is characteristic present? Observation site ______						Subject Sample Type 2 Is characteristic present? Observation site ______					
	Yes (A) (check one)	No (C)		Yes (A) (check one)	No (C)		Yes (B) (check one)	No (D)		Yes (B) (check one)	No (D)
S1			S 26			S 1			S 26		
S2			S 27			S 2			S 27		
S3			S 28			S 3			S 28		
S4			S 29			S 4			S 29		
S5			S 30			S 5			S 30		
S6			S 31			S 6			S 31		
S7			S 32			S 7			S 32		
S8			S 33			S 8			S 33		
S9			S 34			S 9			S 34		
S10			S 35			S 10			S 35		
S11			S 36			S 11			S 36		
S12			S 37			S 12			S 37		
S13			S 38			S 13			S 38		
S14			S 39			S 14			S 39		
S15			S 40			S 15			S 40		
S16			S 41			S 16			S 41		
S17			S 42			S 17			S 42		
S18			S 43			S 18			S 43		
S19			S 44			S 19			S 44		
S20			S 45			S 20			S 45		
S21			S 46			S 21			S 46		
S22			S 47			S 22			S 47		
S23			S 48			S 23			S 48		
S24			S 49			S 24			S 49		
S25			S 50			S 25			S 50		

Total A = ☐ B = ☐ C = ☐ D = ☐

		Sample Type 1	Sample Type 2
Characteristic present?	Yes	*A*	*B*
	No	*C*	*D*

To compute chi-square, transfer numbers to appropriate Boxes in Appendix B.

Comments:

Research Example

Status of Frustrator as an Inhibitor of Horn-Honking Responses[1]

ANTHONY N. DOOB AND ALAN E. GROSS

A. Introduction

Subjects may consciously attempt to present themselves in a favorable manner, they may cooperate with the experimenter or interviewer, and their reactions may be affected by the measurement process itself. In reviewing a number of such problems, Webb *et al.* (6, pp. 13–27) point out that some of these sources of contamination can be avoided when field data are collected from people who are unaware that they are subjects participating in an experiment. Although field procedures can reduce demand and reactivity effects, experimental manipulations outside of the laboratory may gain realism at the expense of control. The study reported here is an attempt to investigate unobtrusively some effects of frustration in a naturalistic setting without sacrificing experimental control.

SOURCE. *The Journal of Social Psychology*, 1968, **76**, 213-218.

[1] We wish to thank Tina Fox and Mike Rosenberg, the observers in the field experiment, and Lorraine Soderstrum of Foothill College, Los Altos Hills, California, who made her class available for the questionnaire experiment. The first author was supported by a Public Health Service Predoctoral Fellowship.

Modern automobile traffic frequently creates situations which closely resemble classical formulations of how frustration is instigated. One such instance occurs when one car blocks another at a signal-controlled intersection. Unlike many traffic frustrations, this situation provides a clearly identifiable frustrator and a fairly typical response for the blocked driver: sounding his horn. Horn honking may function instrumentally to remove the offending driver and emotionally to reduce tension. Both kinds of honks may be considered aggressive, especially if they are intended to make the frustrator uncomfortable by bombarding him with unpleasant stimuli.

One factor that is likely to affect aggressive responses is the status of the frustrator (2,3). The higher a person's status, the more likely it is he will have power to exercise sanctions, and although it is improbable that a high status driver would seek vengeance against a

honker, fear of retaliation may generalize from other situations where aggression against superiors has been punished.

Aggression is not the only kind of social response that may be affected by status. High status may inhibit the initiation of any social response, even a simple informational signal. Although it is difficult in the present study to distinguish informational from aggressive motivation, it is hypothesized that a high status frustrator will generally inhibit horn honking.

B. Method

One of two automobiles, a new luxury model or an older car, was driven up to a signal controlled intersection and stopped. The driver was instructed to remain stopped after the signal had changed to green until 15 seconds had elapsed, or until the driver of the car immediately behind honked his horn twice. Subjects were the 82 drivers, 26 women and 56 men, whose progress was blocked by the experimental car. The experiment was run from 10:30 A.M. to 5:30 P.M. on a Sunday, in order to avoid heavy weekday traffic.

1. STATUS MANIPULATION

A black 1966 Chrysler Crown Imperial hardtop which had been washed and polished was selected as the high status car.[2] Two low status cars were used: a rusty 1954 Ford station wagon and an unobtrusive gray 1961 Rambler sedan. The Rambler was substituted at noon because it was felt that subjects might reasonably attribute the Ford's failure to move to mechanical breakdown. Responses to these two cars did not turn out to be different, and the data for the two low status cars were combined.

2. LOCATION

Six intersections in Palo Alto and Menlo Park, California, were selected according to these criteria: (*a*) a red light sufficiently long to insure that a high proportion of potential subjects would come to a complete stop behind the experimental car before the signal changed to green, (*b*) relatively light traffic so that only one car, the subject's, was likely to pull up behind the experimental car, and (*c*) a narrow street so that it would be difficult for the subject to drive around the car blocking him. Approximately equal numbers of high and low status trials were run at each intersection.

3. PROCEDURE

By timing the signal cycle, the driver of the experimental car usually managed to arrive at the intersection just as the light facing him was turning red. If at least one other car had come to a complete stop behind the experimental car

[2] We have labeled this operation a "status manipulation" because a large expensive car is frequently associated with wealth, power, and other qualities which are commonly regarded as comprising high status. However, it could be argued that Chrysler is potentially inhibiting not because it is a status symbol, but because of some other less plausible attribute (e.g., physical size).

before the signal had turned green, a trial was counted, and when the light changed,an observer started two stop watches and a tape recorder. Observers were usually stationed in a car parked close to the intersection, but when this was not feasible, they were concealed from view in the back seat of the experimental car. High and low status trials were run simultaneously at different intersections, and the two driver-observer teams switched cars periodically during the day. Drivers wore a plaid sport jacket and white shirt while driving the Chrysler, and an old khaki jacket while driving the older car.

a. Dependent Measures

At the end of each trial, the observer noted whether the subject had honked once, twice, or not at all. Latency of each honk and estimated length of each honk were recorded and later double-checked against tape recordings.

b. Subject Characteristics

Immediately after each trial, the observer took down the year, make, and model of the subject's car. Sex and estimated age of driver, number of passengers, and number of cars behind the experimental car when the signal changed were also recorded.

C. Results and Discussion

Eight subjects, all men, were eliminated from the analysis for the following reasons: four cars in the low status condition and one in the high status condition went around the experimental car; on one trial the driver of the experimental car left the intersection early; and two cars in the low status condition, instead of honking, hit the back bumper of the experimental car, and the driver did not wish to wait for a honk. This left 38 subjects in the low status condition and 36 in the high status condition.

Although the drivers of the experimental cars usually waited for 15 seconds, two of the lights used in the experiment were green for only 12 seconds; therefore 12 seconds was used as a cutoff for all data. There were no differences attributable to drivers or intersections.

The clearest way of looking at the results is in terms of the percentage in each condition that honked at least once in 12 seconds. In the low status condition 84 per cent of the subjects honked at least once, whereas in the high status condition, only 50 per cent of the subjects honked ($\chi^2 = 8.37$, $df = 1$, $p < .01$).[3] Another way of looking at this finding is in terms of the latency of the first honk. When no honks are counted as a latency of 12 seconds, it can be seen in Table 1 that the average latency for the new car was longer for both sexes. ($F = 10.71, p < .01$).

Thus, it is quite clear that status had an inhibitory effect on honking even once. It could be argued that status would have even greater inhibitory effects on more aggressive honking. Although one honk can be considered a polite way of calling attention to the

[3] Notations such as $p < .01$ are explained in the Introduction, pp. 15-16.

green light, it is possible that subjects felt that a second honk would be interpreted as aggression.[4]

Table 1

Field Experiment (Mean Latency of First Honk in Seconds)

Frustrator	Sex of driver: Male	Female
Low status	6.8 (23)	7.6 (15)
High status	8.5 (25)	10.9 (11)

Note: Numbers in parenthesis indicate the number of subjects.

Forty-seven per cent of the subjects in the low status condition honked twice at the experimental car, as compared to 19 per cent of the subjects in the high status condition ($\chi^2 = 5.26$, $df = 1$, $p < .05$). This difference should be interpreted cautiously because it is confounded with the main result that more people honk generally in the low status condition. Of those who overcame the inhibition to honk at all, 56 per cent in the low status condition and 39 per cent in the high status condition honked a second time, a difference which was not significant. First-honk latencies for honkers were about equal for the two conditions. The overall findings are presented in Table 2.

Table 2

Number of Drivers Honking Zero, One, and Two Times

Frustrator	Honking in 12 seconds: Never	Once	Twice
Low status	6	14	18
High status	18	11	7

Note: Overall $\chi^2 = 11.14, p < .01$.

[4] Series of honks separated by intervals of less than one second were counted as a single honk.

Sex of driver was the only other measure that was a good predictor of honking behavior. In both conditions men tended to honk faster than women ($F = 4.49, p < .05$). The interaction of status and sex did not approach significance ($F = 1.17$). These data are consistent with laboratory findings (1) that men tend to aggress more than women.

Most experiments designed to study the effects of frustration have been carried out in the laboratory or the classroom, and many of these have employed written materials (2,5).

It is undoubtedly much easier to use questionnaires, and if they produce the same results as field experiments, then in the interest of economy, they would have great advantage over naturalistic experiments. However, over 30 years ago, LaPiere warned that reactions to such instruments "may indicate what the responder would actually do when confronted with the situation symbolized in the question, but there is no assurance that it will" (4, p.236).

In order to investigate this relationship between actual and predicted behavior, an attempt was made to replicate the present study as a questionnaire experiment. Obviously, the most appropriate sample to use would be one comprised of motorists sampled in the same way that the original drivers were sampled. Because this was not practicable, a questionnaire experiment was administered in a junior college classroom.

Subjects were 57 students in an introductory psychology class. Two forms of the critical item were included as the first of three traffic situations on a one-

page questionnaire: "You are stopped at a traffic light behind a black 1966 Chrysler (gray 1961 Rambler). The light turns green and for no apparent reason the driver does not go on. Would you honk at him?" If subjects indicated that they would honk, they were then asked to indicate on a scale from one to 14 seconds how long they would wait before honking. Forms were alternated so that approximately equal numbers of subjects received the Chrysler and Rambler versions. Verbal instructions strongly emphasized that subjects were to answer according to what they actually thought they would do in such a situation. No personal information other than sex, age, and whether or not they were licensed to drive was required.

After the questionnaire had been collected, the class was informed that different kinds of cars had been used for the horn-honking item. The experimenter then asked subjects to raise their hands when they heard the name of the car that appeared in the first item of their questionnaire. All subjects were able to select the correct name from a list of four makes which was read.

One subject (a female in the high status condition) failed to mark the honk latency scale, and another subject in the same condition indicated that she would go around the blocking car. Both of these subjects were eliminated from the analysis, leaving 27 in the high status condition and 28 in the low status condition. The results were analyzed in the same manner as the latency data from the field experiment. Means for each condition broken down by sex are presented in Table 3. Males reported that they thought that they would honk considerably sooner at the Chrysler than at the Rambler, whereas this was slightly reversed for females (interaction of sex and status $F = 4.97, p < .05$). Eleven subjects, six males in the low status condition and five females in the high status condition indicated that they would not honk within 12 seconds.

Table 3
Questionnaire Experiment (Mean Latency of Honking in Seconds)

	Sex of subject	
Frustrator	Male	Female
Low status	9.1 (18)	8.2 (10)
High status	5.5 (13)	9.2 (14)

Note: Numbers in parentheses indicate the number of subjects.

It is clear that the behavior reported on the questionnaire is different from the behavior actually observed in the field. The age difference in the samples may account for this disparity. Median estimated age of subjects in the field was 38, compared to a median age of 22 in the classroom. In order to check the possibility that younger males would indeed honk faster at the high status car, the field data were reanalyzed by age. The results for younger males, estimated ages 16 to 30, fit the general pattern of the field results and differed from the results of the classroom experiment. In the field, young males honked sooner at the Rambler than at the Chrysler ($t = 2.74, df = 11, p < .02$).

Unfortunately, because these two studies differed in both sample and method, it is impossible to conclude that the differences are due to differences in the method of collecting data. However, it is clear that questionnaire data obtained from this often used population of subjects do not always correspond to what goes on in the real world.

References

1. Buss, A. H. Instrumentality of aggression, feedback, and frustration as determinants of physical aggression. *J. of Personal. & Soc. Psychol.,* 1966, **3** 153–162.
2. Cohen, A. R. Social norms, arbitrariness of frustration, and status of the agent in the frustration-aggression hypothesis. *J. Abn. & Soc. Psychol.*, 1955, **51**, 222–226.
3. Hokanson, J. E., & Burgess, M. The effects of status, type of frustration and aggression on vascular processes. *J. Abn. & Soc. Psychol.*, 1962, **65**, 232–237.
4. LaPiere, R. T. Attitudes vs. actions. *Social Forces*, 1934, **13**, 230–237.
5. Pastore, N. The role of arbitrariness in the frustration-aggression hypothesis. *J. Abn. & Soc. Psychol.*, 1952, **47**, 728–731.
6. Webb, E. J., Campbell, D. T., Schwartz, R. D. & Sechrest, L. Unobtrusive Measures: Nonreactive Research in the Social Sciences. Chicago, Ill.: Rand McNally, 1966.

Exercise 4

Social Reinforcement and Verbal Conditioning

Reinforcement is a technique that is commonly used to increase the likelihood that a certain behavior will occur. Any event is defined as "reinforcing" if, when it follows a behavior, it increases the probability that that behavior will occur again. Thus reinforcers are generally positively valued events, usually described as "good" or "pleasurable," that have the effect of encouraging the behavior which they follow. *Social* reinforcement techniques are used to control the behavior of others by giving social approval after the subject emits the desired behavior. Verbal and nonverbal means of communicating social approval both can be used to influence responses in a specified direction. Verbal statements such as "uh-huh," "good," "fine," as well as nonverbal signs such as smiling, nodding, or leaning toward another, are all common techniques for manipulating or reinforcing another person's behavior. When this type of conditioning procedure is applied successfully to alter an individual's verbal behavior, it is termed *verbal conditioning*.

One of the most intriguing aspects of the study of verbal conditioning is the possibility that such alterations in the frequency of certain verbal behavior may occur without the subject's awareness that his behavior is being changed. This kind of learning without intention or consent has been labeled "nonconscious," or "learning without insight."

Initially, it was thought that verbal behavior was controllable without the subject's awareness of the manipulation (Krasner, 1958), but subsequent researchers challenged this conclusion. For example, Spielberger (1962) presented evidence that conditioning occurred only when the subject was aware that the critical response was followed by a reinforcement. He argued

that verbal conditioning without awareness had not been demonstrated, and a lively controversy began. Although no one challenged Spielberger's findings that subject awareness of the reinforcer was related to the success of the conditioning procedure, several questions have been raised regarding interpretation of his results. First, it was possible that subjects were not aware prior to the interview, but that Spielberger's interview method for assessing awareness actually suggested the reinforcement process to the subjects. Second, subject awareness could be the result, not the cause, of his learning. Thus, it was argued that alteration of verbal behavior produced awareness, not vice versa.

In this exercise you will attempt to influence another person's statements by giving him verbal approval ("good," "fine," etc.) whenever he emits one class of verbal responses instead of another. If your social reinforcement is effective, the kinds of statements that are followed by your approving comments should increase in frequency. After you attempt to reinforce your subject for saying things you want him to say, you will interview him to determine to what degree he was aware of your attempts to condition his verbal behavior.

Procedure

Before recruiting your subjects, you should prepare 40 3 × 5 inch index cards. On each card list six pronouns (I, he, she, we, you, they) and a common verb as illustrated in Fig. 1. These cards will be the stimulus material for the experiment.

I He She We You They	Swim

Figure 1. Sample stimulus card.

You should recruit sixteen subjects, eight of whom will receive your approval ("good," "uh-huh") whenever they use either the pronoun "I" or "we" in constructing a sentence. These eight subjects comprise the experimental group. The remaining eight subjects should not be given any reinforcement. Subjects should be tested one at a time without anyone else present. It is important to note that social reinforcements should be carefully controlled in this experiment. Only give reinforcements following sentences beginning with the critical words "I" and "we" and do not show

signs of approval on any other occasion. The exact instructions to be given each subject are provided on the Experimenter's Script and Instructions (p. 83).

Data Analysis

Choose a partner who has also completed this exercise and exchange with him the data obtained on the "awareness" questionnaire. The purpose of trading data is to allow "blind" judgments of the subject's awareness. If you were aware of how a particular subject performed during the verbal conditioning task, such knowledge could influence your ratings of that subject's insight into the experimental procedures. Thus each of you can serve as a "blind" rater for the other. Your task is to categorize each subject as either "aware" or "unaware." Aware subjects are those who specify, at some point in their interview, that the experimenter was giving rewards following each "I" and "we" response. After each subject is classified as aware or unaware, return the data to your partner, and code your subjects for awareness in the space provided at the bottom of Data Sheets 1 and 2.

Experimental-Control Comparison

Next compare the control and experimental groups for verbal conditioning on Data Sheets 1 and 2. First add the number of "I" and "we" responses made for each subject during his 40 trials. Do this separately for experimental and comparison groups (on Data Sheets 1 and 2). Then determine the mean number of responses for the subjects in each of the two groups. That is, total the number of responses made by the subjects in each group and divide by the number of subjects ($N = 8$). These two means can then be compared for significance by use of the t-test (see Appendix A). Were there significant differences between the experimental and control groups?

Learning Curves. Next you will determine if subjects in the experimental group demonstrated learning. That is, did the frequency of "I" and "we" responses increase as a function of the number of trials and reinforcements? Did "I" and "we" responses increase from the first to the last trials? Plot the curves for the experimental and the control group. To do this sum the number of critical responses made by subjects within each group (Experimental and Comparison) for the first ten trials (cards) and divide by the number of subjects; do the same for the next block of ten trials; then the next block; and finally the last block of trials. Space for computing block means is provided on Data Sheets 1 and 2. Plot the mean number of responses given by members in the experimental and comparison groups on the various trial blocks. An example is given in Fig. 2.

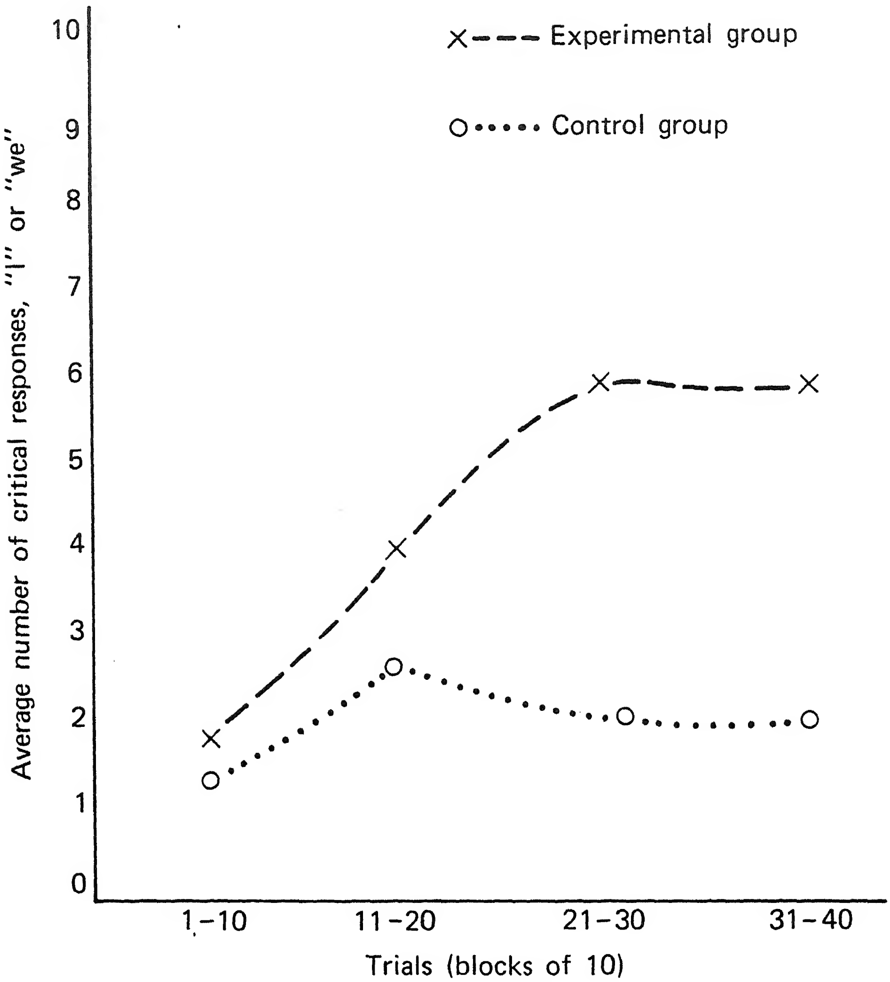


Figure 2.

Next you will determine if this group learning curve actually reflects the performance of individual members within each group. To determine this, plot the number of responses given by *each* experimental subject and the previously computed "average" curve on the same piece of graph paper. An example of how such a plotting might appear is given in Fig. 3.

Does the average learning curve for the Experimental group accurately represent the performance of most of the individual subjects you ran?

The Effect of Awareness. To determine the relationship between awareness and verbal performance, group your experimental subjects into those who were rated as being aware and those who were not. Did the aware subjects emit more critical responses ("I" and "we") than subjects who

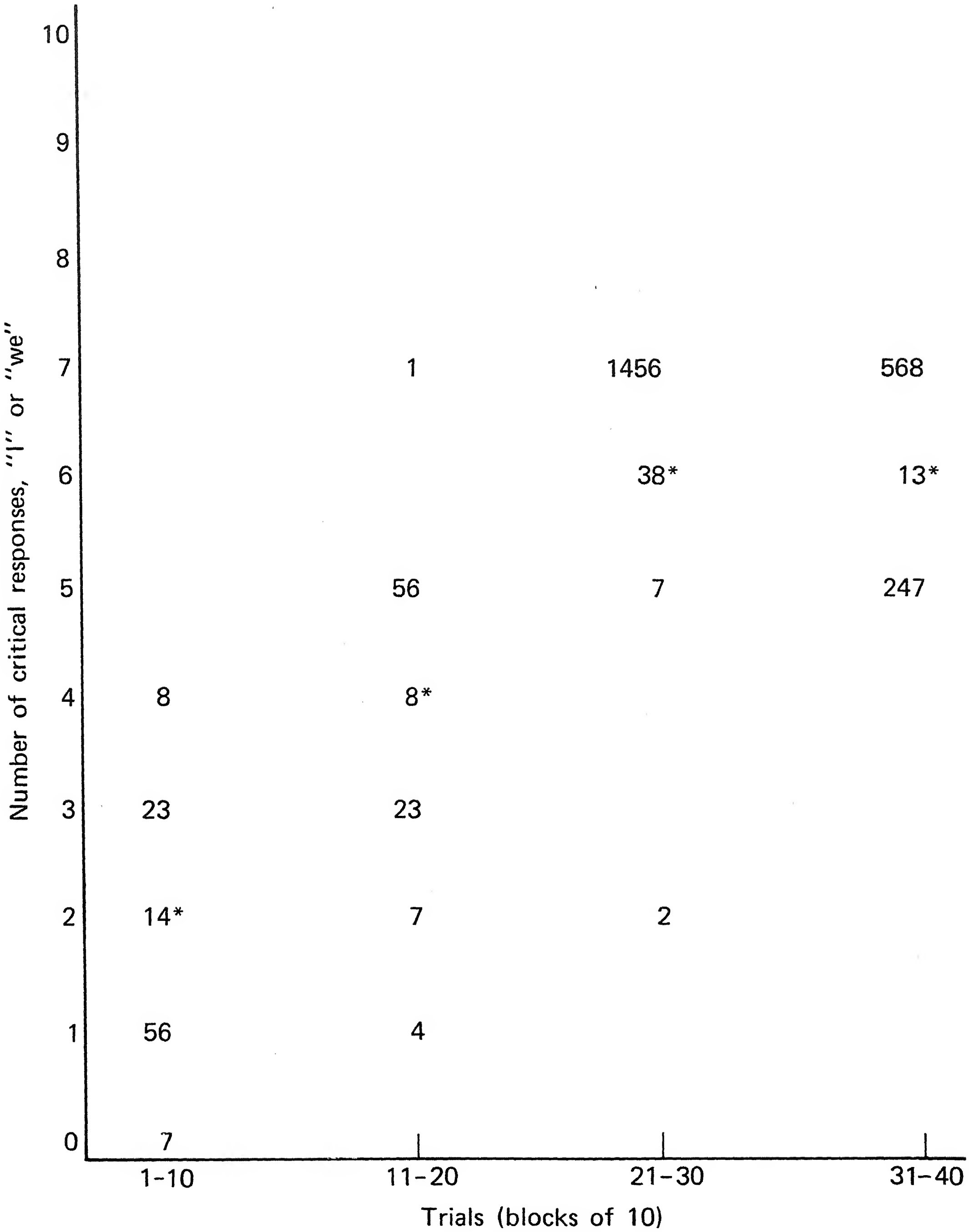


Figure 3. Numbers represent individual scores for hypothetical experimental subjects 1–8. The asterisk indicates average score for experimental group.

received reinforcements but who did not become aware? Plot learning curves for these two groups of subjects. Does average performance for aware subjects resemble that of the unaware but reinforced participants? Do the individual learning curves differ for unaware and aware subjects, and if so, can you detect on what trial block the aware subjects became aware?

Last, but not least, what ethical implications do you see for psychologists who produce change without the subject's awareness? Is this an invasion of privacy? Are subjects being "brainwashed?" Do the potential benefits from such research offset these risks? Are the results from such experiments likely to advance science or human welfare?

A NATURALISTIC (FIELD) VERSION OF VERBAL CONDITIONING (OPTIONAL)

Engage a friend in a conversation about something he is interested in. Select a topic or class of sentences that you will reinforce. For example you may reinforce any sentence beginning with or including the pronouns "I," "we," "mine," or "myself" (self-reference statements). During the first five minutes of the conversation, simply count the number of such statements. This will give you a base line of normal or operant behavior. Then during the next five minutes, administer reinforcements after each statement in the selected category. During the conversation, you can answer direct questions, but otherwise your comments should be extremely brief. Let the other person talk. Compare the number of reinforced statements in the first five minutes (operant period) with the second five minutes (acquisition period). If there was an increase in self-reference statements, was this effect a result of your reinforcements? Can you think of other rival interpretations for this effect? A classic study by Verplanck, which demonstrates social reinforcement of self-reference statements in a field setting, is reported in the following Research Example on page 91.

References

Bandura, A. *Principles of behavior modification.* New York: Holt, Rinehart and Winston, 1969.

Chapter 9 of this book presents an excellent review of the research on learning without awareness and the controversy that has surrounded it.

Eriksen, C. W. (Ed.) *Behavior and awareness: The symposium of research and interpretation.* Durham, N. C.: Duke University Press, 1962.

This edited book presents major theories concerning the role of awareness on the learning process. Various interpretations of the interaction of thoughts and learning are presented by the major theorists in this area of investigation.

Krasner, L. Studies of the conditioning of verbal behavior. *Psychological Bulletin*, 1958, **55**, 148–170.

This article reviews the early work in the area of verbal conditioning; it served as an important stimulus in increasing the interest in the area of "learning without awareness."

Spielberger, C. D. The role of awareness in verbal conditioning. *Journal of Personality*, 1962, **30**, 73–101.

This study casts doubt on the presumption that verbal conditioning took place without subject awareness.

Experimenter's Script and Instructions

"I am doing a psychology experiment on linguistic habits. Your task is to take these cards (show first card), and make up a sentence using one of the pronouns on the left side of the card and the verb on the right side. Go through the deck one at a time and simply make up a sentence using a pronoun and the verb. Continue making up sentences until you complete the deck. Okay, go ahead."

As the subjects give their sentences, check those trials in which they use either the "I" or "we" pronouns on Data Sheets 1 and 2. For the experimental subjects be sure to say "good" and "uh-huh" each time the subject uses either an "I" or "we" pronoun. For the remaining eight comparison subjects remain silent throughout the forty trials. Following the task, administer the Awareness Questionnaire on p. 85 and record his answers word for word.

Awareness Questionnaire

After the conditioning procedure (whether successful or not), question the subject to find out if he was aware of the purpose of the exercise. Probe gently starting with very general questions and proceeding to more specific suggestions. A series of probe questions follows:[1]

Did you say the first sentence that came into your mind?

How did you decide what sentence to say?

Were you using some words on the card more than others? Which words? Why?

What do you think the purpose of this experiment was?

Did you think you were supposed to say certain sentences?

What was I doing while you were speaking?

Did you notice that I said (did) anything?

Actually I did nod my head (say hmm-hmm, etc.). Do you remember that?

What did my head nod mean to you?

Can you figure out when or why I nodded my head?

Actually I nodded my head only after you began a sentence with a certain kind of word. Do you notice which kind of word?

Did you ever have the idea that I was nodding my head after you began sentences with "I" or "we"?

[1] Adapted from Charles D. Spielberger "The role of awareness in verbal conditioning" in C. W. Eriksen (Ed.) *Behavior and Awareness*. Durham: Duke University Press, 1962, p. 86.

Data Sheet 1

Checks = Responses "I" + "we"

Trial	Subject Number 1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								
Totals								

Data Sheet 2

Checks = Responses "I" + "we"

Subject Number

Trial	1	2	3	4	5	6	7	8
1								
2								
3								
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18								
19								
20								
21								
22								
23								
24								
25								
26								
27								
28								
29								
30								
31								
32								
33								
34								
35								
36								
37								
38								
39								
40								
Totals								

Research Example

The Control of the Content of Conversation: Reinforcement of Statements of Opinion[1]

WILLIAM S. VERPLANCK

Some kinds of human behavior have seemed to be resistant to experimental investigation because of both their complexity and their apparent variability. One such class includes the commonplace activities of people—for example, whatever the reader was doing just before he picked up this journal. Perhaps talking to someone.

This paper describes the successful experimental application of some principles of operant conditioning in this area; specifically to conversation between two people. The experimental procedure is based on two assumptions. (*a*) Apparently heterogeneous human verbal behavior falls into comparatively simple operant response classes; hence, any one is susceptible to conditioning. The class of verbal behavior chosen is the *stating of opinions.* (*b*) Classes of environmental events can be isolated that have the property of altering any behavior on which their occurrence has depended, i.e., some events are reinforcing stimuli. Specifically, under our conditions, statements of *agreement* or *paraphrase* are hypothesized to be reinforcing stimuli for the verbal behavior of a speaker. According to these assumptions, if someone agrees with every opinion of a speaker, the speaker should show a sharp increase in his rate of stating opinions. The *stating of opinions* has been conditioned.

Since it is both interesting and important to obtain changes in behavior that correspond to those termed conditioning when the subject is not aware that he is "being conditioned" (or, indeed, that his behavior is being manipulated in any way) the present experiments were conducted under conditions in which the "insight" was extremely unlikely.

SOURCE. *Journal of Abnormal and Social Psychology*, 1955, **51**, 668–676.

[1]The first experiments on this subject were carried out by Mr. Ronald M. Dworkin, as an experimental project in an undergraduate course. His exploratory results were indispensable in setting up the procedures followed in this experiment.

Method

GENERAL PLAN OF THE EXPERIMENT

The experiment was carried out in a series of ordinary conversations between

Table 1
Treatments Followed by Experimenters

N	First 10 Minutes	Second 10 Minutes	Third 10 Minutes
5	O—Measure operant level	A—Reinforce each opinion-statement by agreement	D—Extinguish by disagreeing with each opinion-statement
2	O—Measure operant level	A—Reinforce each opinion-statement by agreement	E—Extinguish by failing to respond to any statement of *S* (silence)
6	O—Measure operant level	P—Reinforce each opinion-statement by paraphrase	D—Extinguish by disagreeing with each opinion-statement
4	O—Measure operant level	P—Reinforce each opinion-statement by paraphrase	E—Extinguish by failing to respond to any statement of *S* (silence)
7	A_1—Reinforce each opinion-statement by agreement	E—Extinguish by failing to respond to any statement of *S* (silence)	A_2—Reinforce each opinion-statement by agreement

"stating opinions" by ticking off the total number of statements and the num- two people, the subject (*S*) who was not informed in any way that he was taking part in an experiment, and the experimenter (*E*). The conversations lasted at least a half-hour which was divided into three 10 minute periods.

During the first 10-minute period, once the conversation was under way, *E* did not reinforce any statement made by *S*, but determined his operant level of ber of opinion-statements made by *S* in successive one-minute intervals. This treatment for the first 10-minute period is labeled O in the first column of Table 1.

In the second 10 minutes, every opinion-statement *S* made was recorded by *E* and reinforced. For two groups, *E* agreed with every opinion-statement by saying: "Yes, you're right," "That's so," or the like, or by nodding and smiling affirmation if he could not interrupt. This treatment is labeled A, for agreement, in the second column of Table 1. For two other groups, *E* reinforced by repeating back to *S* in paraphrase each opinion-statement that *S* made (labeled P in the second column of Table 1).

In the third 10-minute period, the *E*s attempted to extinguish the opinion-statements of two groups by withdrawing *all* reinforcement, that is, by failing to respond (labeled E for extinguish in the third column of Table 1) in any way to *S's* speech, and of two other groups by disagreeing with each opinion stated (labeled D in the third column of Table 1).

The design of the experiment is depicted in Table 1. Of the four O-groups of the first period, two become groups in which reinforcement came by agreement (A-groups) in the second period, and two became groups in which reinforcement

came by paraphase (P-groups). In the third period, one of the A-groups was extinguished by disagreement (D-group) and one by *E's* silence (E-group). A similar division was made for the P-groups. Thus, each of the four groups can be designated by the combination of treatments provided in the three consecutive periods of conversation.

In the fifth, control group ($A_1 EA_2$), run to insure that any changes in *S's* rate of stating opinions could not be attributed simply to the passage of time during the experiment, *E* reinforced by agreement *S's* opinion-statements in the first and third 10-minute periods, and withdrew all reinforcement during the second period.

During the first (O) period for the first four groups, and the E period for the fifth group, *E* asked a "neutral" question ("What did you say?") if *S's* rate of speaking showed signs of declining. Few such were necessary.

EXPERIMENTAL SITUATION

The *E's* performed the experiment when and where they could, restricted by only three criteria: (*a*) that only two persons be present, (*b*) that there be a clock, and the paper and pencil required for recording, and (*c*) that enough time be available to both *S* and *E* for them to talk for at least a half hour. The *E's* did not suggest to *S's* at any time that an experiment was being carried on, and in the rare cases in which an *S* showed signs of suspicion that this was not an ordinary conversation the experiment was terminated (although the conversation was carried on).

Seventeen *S*s were run in student living quarters, two in restaurants, two in private homes, and one each in a hospital ward, in a public lounge, and over the telephone. In one experiment, contrary to instructions, a third (but uninformed) person was present.

The topics of conversation ranged from the trivial to the "intellectual" and included dates, vacations, Marxism, theory of music, man's need for religion, architecture, Liberace.

EXPERIMENTERS

Seventeen members of a course[2] in the Psychology of Learning served as *E*s. Twelve were Harvard undergraduates, two were Radcliffe undergraduates, and three (two women and one man) were students in the Graduate School of Education. All the experimenters had had extensive experience in the techniques of conditioning bar-pressing in the rat, and of conditioning chin-tapping in the human (3). Of the 17 students who undertook the experiment, all were able to collect one or two sets of data as the design demanded.

SUBJECTS AND EXPERIMENTAL GROUPS

Of the 20 men and four women who served as *S*s, 13 were described by the *E*s

[2] An experiment of this sort very probably could not be successfully performed *de novo* in a laboratory situation suitably equipped for tape-recording and concealed observation. The present strategy was dictated by the need to determine whether positive results could be obtained in conversations on a variety of topics, carried on in a wide variety of situations, and especially in a situation in which it was most unlikely that *S* would suspect that an experiment was being carried on.

as friends, seven as roommates, one a date, one an uncle, and one a total stranger. In all but four conversations, *S* and *E* were of the same sex. All but six *S*s were of college age; of these six, four were in the thirties, and two were 55 and 60, respectively.

These *S*s were distributed over the four experimental groups as follows: OAD, 5; OPD, 6; OPE, 2; OAE, 4; $A_1 EA_2$, 7.

There were 20 students in the class, and the design called for *N*'s of 5 and 10, but 3 students reported that they were unable to undertake the experiment,[3] and of the 17 *E*s, one placed himself in the wrong group.

THE RESPONSE CONDITIONED

The response selected for reinforcement was the uttering by *S* of a statement or "sentence" beginning: "I think . . .," "I believe . . .," "It seems to me," "I feel," and the like. The *E*s were instructed to be conservative in classifying a statement as an opinion, and to do so only if one or another such qualifying phrase began the statement. (*E*s were aware that the experiment was designed to investigate *S*s' behavior, and not their own.) No attempt was made to define what constituted a statement or a "sentence" except that *E* should not expect grammatical sentences (1). These instructions proved adequate; no *E* had difficulty in counting such units of verbal behavior, although doubtless many speech units counted would not parse.

[3] That three *E*s found themselves unable to undertake the experiment is in itself interesting. A fourth resorted to the telephone, with good results.

REINFORCING STIMULI

Two classes of reinforcing stimuli were used by the *E*s. The first was *agreement* (A), defined as the experimenter saying "You're right," "I agree," "That's so," or the like, nodding the head, smiling (where *E* did not want to interrupt). The second was repeating back to *S* in paraphrase (P) what he had just said. No further attempt was made to specify paraphrasing. *Extinction* was carried out in one of two ways. In some groups *E* simply refrained from responding in any way to a statement by *S*(E) and in others, he disagreed (D) with each opinion-statement.

The *E*s did not speak, except to reinforce, to disagree, or to "prime" *S* with a question during operant-level determination. They contributed nothing new to the conversation.

RECORDING

A clock, or watch with sweep-second hand, a pencil, and something to write on were necessary for the recording. One *E* was able to record the whole conversation on a tape-recorder. The *E*s ticked off each statement occurring in successive one-minute intervals by making a series of doodles incorporationg marks, or by making marks on the margin or text of a book or magazine. Different marks were used for opinions and other statements. Recording proved inconspicuous, and in only one or two cases did an *E* have to terminate an experiment because *S* seemed to notice his recording.

Although problems arose occasionally, *E*s by and large had no difficulty in

arriving at and maintaining a criterion for a "sentence" or "statement," i.e., for the unit of speech that they counted, and for the subclass, statement of opinion.

The criteria varied from experimenter to experimenter, in that the rates of speaking of two subjects reported by the same *E*s are correlated, and the reported rates are a function not only of the subject's rate of speaking, and of *E*'s rate of speaking in reinforcing, but also of the criterion for "statements" adopted by *E*.

In only one case did an *S* comment on *E*'s recording: during extinction he asked *E* what she was doodling, and was satisfied when she showed him her scribbles. The *E*s also noted *S*'s general behavior during extinction, and the mode of termination of the experiment.

EXECUTION

In a few cases, the experiment was begun, and then terminated by phone calls, third persons entering the room, or because *E* feared that *S* had noticed that he was recording. All the experiments completed are reported in this paper, except one from group A_1EA_2, whose data could not be accurately transcribed. Under questioning, no experimenter reported that he terminated the experiment because results did not seem satisfactory to him.

Two *E*s carried out operant-level determination for only 9 minutes, and one went overtime. Four went overtime during reinforcement. The greatest variability appeared during extinction; seven *S*s failed to continue talking for 10 minutes following the beginning of disagreement, or of nonreinforcement, either leaving the room or falling into silence. Eight *E*s carried on the conversation past the 10-minute minimum extinction period. Since *E*s were not consistent in continuing to record or to converse past this time, data are reported only on the first 10 minutes.

In summary, the experiment is designed to determine whether a person, in conversation with another person, can manipulate the second person's conversation by agreeing or disagreeing, or by paraphrasing. The experimenter himself, it should be noted, contributes nothing new to the content of the conversation.

Results

AWARENESS

No *S* ever gave any evidence that he was "aware" that he was serving as a subject in an experiment, that his behavior was being deliberately manipulated and recorded, or that he recognized that there was anything peculiar about the conversation. The only qualification that must be made is this: during extinction, some *S*s got angry at *E* and commented on his disagreeableness, or noted his "lack of interest," and during reconditioning one member of group A_1EA_2 gave *E* "queer, searching glances," perhaps because of the opinions that *E* was now agreeing with. These changes of behavior are consistent with those found in other situations when *S* is undergoing extinction (3).

Conditioning is demonstrated if the

Table 2
Median and Ranges for Each 10-Minute Period

	10-Minute Period	Groups OAE, OAD, OPE, OPD combined			Group $A_1 EA_2$		
		Proc.	Median	Range	Proc.	Median	Range
Rate	1st	op	5.3	2.2-12.8	cond	7.1	2.4-14.0
(statements/minute)	2nd	cond	5.7	3.2-17.1	ext	6.3	1.9-11.0
	3rd	ext	5.2	1.4-12.8	recond	5.8	2.9-14.5
Relative frequency	1st	op	0.320	.012-.655	cond	0.574	.208-.653
of opinion - statements	2nd	cond	0.558	.071-.702	ext	0.302	.094-.526
	3rd	ext	0.333	.048-.643	recond	0.603	.267-.699

appropriate changes appear in the rate of speaking opinion-statements as a function of the conditions of reinforcement. When reinforcement is given, the rate must increase; when it is withdrawn, the rate must decrease.

Distributions were made of the number of opinion-statements (N_{opin}) and of all statements (N_{all}), and their cumulative values (CN_{opin} and CN_{all}) for each minute of the three experimental periods. From the latter, mean rates of making statements were computed. Relative frequencies of opinions ($RF_{opin} = CN_{opin}/CN_{all}$) were determined for each S for each period.

RATES

The rates of making statements (CN_{all}/t) showed no significant changes as a function of reinforcement. Table 2 gives, in the upper portion, data on the distribution of these rates for each interval. Several nonparametric tests for significance of difference were made, and none showed that the null hypothesis (no difference as a function of period, manipulation, or group) could be rejected. The "priming" of *S* by means of the question, "What did you say?," seems to maintain the rates in the operant periods, and in the extinction period of group A_1EA_2 although decreases in rate may be obscured by the fact that *E* is saying little during these times. The rank-order correlation of operant-level rates of speech obtained on two *S*s by the same *E*s was 0.65 ($N = 14$). This figure includes data to be reported elsewhere but obtained under the same conditions.

RELATIVE FREQUENCY OF OPINIONS

Table 2 (lower portion) presents the medians and ranges of the distributions of RF_{opin} for each period. Each of the 24 *S*s showed an increase in his relative frequency of opinion during the reinforcement period over his operant level, or (for group A_1EA_2) over his preceding extinction period. The probability that this result would have been obtained if there had been no effect of the experimental variable is $(1/2)^{24}$. Twenty-one of the 24 showed a *reduced* RF_{opin} in the extinction or disagreement period below that of the preceding period of reinforcement. The probability that fewer than four *S*s would not change in the absence of an effect of the experimental variable is $1.1\ (1/2)^{13}$. Signed rank tests (4) of the significance of the differences yield *p* values[4] well below .01.

The magnitude of the effects can be evaluated by determining two ratios for each *S*: (*a*) that of RF_{opin} obtained during conditioning to RF_{opin} of the operant level or (for group A_1EA_2) RF_{opin} in reconditioning to RF_{opin} in the preceding extinction period, and (*b*) of RF_{opin} during the extinction period to RF_{opin} during the preceding conditioning period. Large values of the former of these ratios are possible only when the operant level RF_{opin} is low. Table 3 presents the mean, median, and range of these values for groupings of the 24 *S*s based on the methods of reinforcement and extinction.

An evaluation was made of the rela-

[4] Notations concerning *p* values are explained in the Introduction, pp. 15–16.

Table 3
Means, Medians, and Ranges of Ratio-Index of Changes in Relative Frequency of Opinion-Statements

Groups Combined	RF Ratios in Distribution	*N*	Mean	Median	Range
	A. Conditioning Effect (No effect: Ratio-Index = 1.00)				
OAD, OAE	A/O	9	2.27	1.76	1.50- 5.70
$A_1 EA_2$	A_2/E	7	2.29	2.17	1.09- 4.32
OAD, OAE, $A_1 EA_2$	A/O, A_2/E	16	2.28	1.85	1.09- 5.70
OPD, OPE	P/O	8	4.23	2.02	1.05-11.47
All	A/O + A_2/E + P/O	24	2.91	1.85	1.05-11.47
	B. Extinction Effect (No effect: Ratio-Index = 1.00)				
OPE, OAE	E/P, E/A	11	0.71	0.70	0.48-0.86
$A_1 EA_2$	E/A_1	7	0.66	0.52	0.45-1.15
OPE, OAE, $A_1 EA_2$	E/P, E/A, E/A_1	18	0.69	0.52	0.45-1.15
OPD, OAD	D/P, D/A	6	0.65	0.62	0.27-1.01
All	E/P, E/A, E/A, D/P, D/A	24	0.67	0.65	0.27-1.15

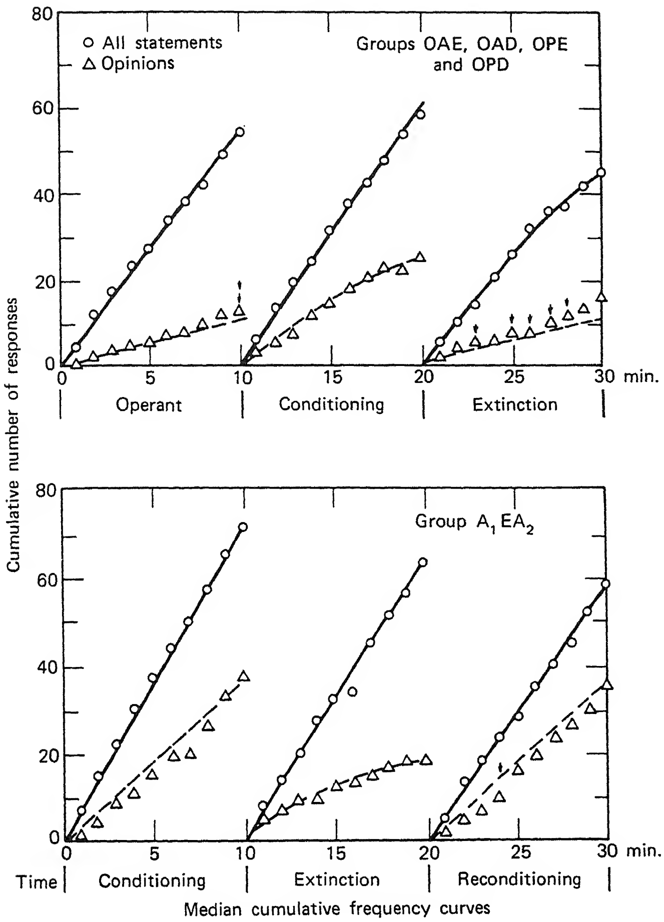


Figure 1. Median cumulative frequency curves of opinion-statements, and of all statements, for each 10-minute period of the experiment

For the upper graphs N = 17, for the lower, N = 7. At each arrow, N on that and successive trials is diminished by one. In the extinction period of the upper graph, each S that dropped out, "had to leave." In the other cases, E discontinued the procedure at the time indicated.

tive effectiveness of agreement and paraphrase in conditioning, and of disagreement and silence in extinction. Fisher's exact test of independence in contingency tables was applied about the medians of Table 3A for groups OAD and OAE taken together versus OPE and OPD, and about the medians of Table 3B for groups OAD and OPD against OAE and OPE. No difference in the number of cases falling above and below the medians was significant at the .05 level, although the difference between agreement and paraphrase is significant between the .05 and .10 levels.

Means and variances were also computed. An F test of the significance of difference in the variances of OAD and

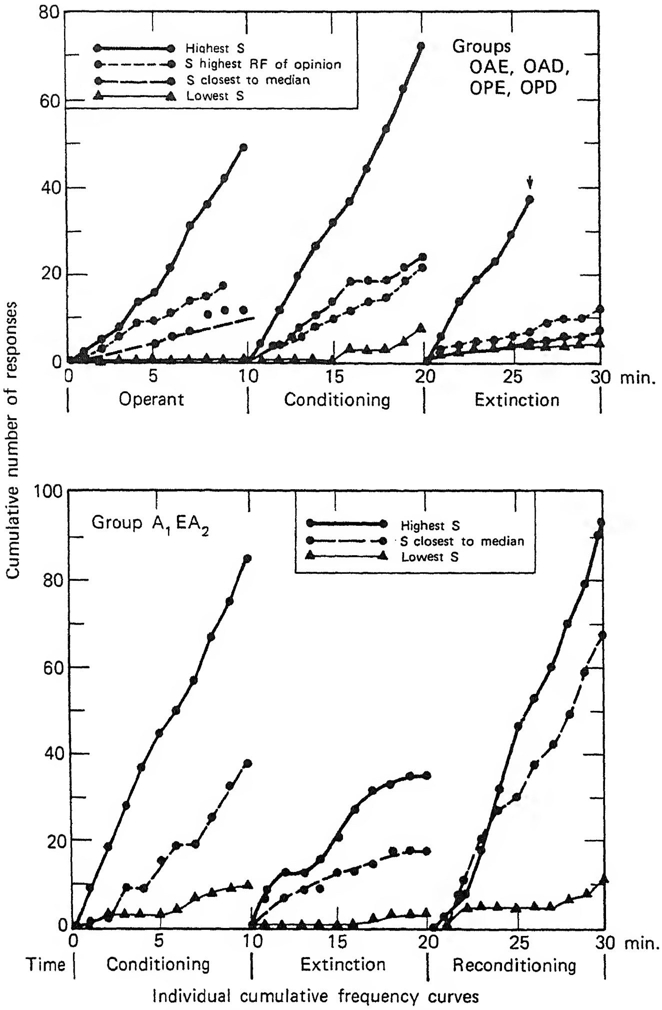


Figure 2. Individual cumulative frequency curves of opinion-statements for each 10-minute period of the experiment, demonstrating the consistency of the effect and its range

OAE and of OPD and OPE gives 8.239 (df = 8, 7), significant at better than the .005 level. Paraphrasing and agreement, although both effective, are not equivalent as reinforcing stimuli; paraphrasing is much more variable in its effectiveness (or perhaps the variety of statements made as paraphrases exceeded those called agreements).

The method of extinction also yielded a significant difference in variance: F = 5.175 (df = 10,5), significant at the .05 level. Despite these differences in variance, group curves were constructed. All four groups were combined without respect to method of reinforcement or extinction. The median N and CN of opinions, and of all sentences, were then

determined for each successive minute of each of the three periods. Figure 1 (see page 99) presents these medians for the groups OAD, OPD, OAE, and OPE, and for group A_1 EA_2.

Figure 2 demonstrates that the median curves are indeed representative. In it are plotted the experimental points obtained during the operant level period from (*a*) the *S* giving a CN_{op} equaling the median, together with the *S*s giving (*b*) the lowest and (*c*) the highest values among the 17 *S*s of the combined groups, and from the corresponding *S*s of group $A_1 EA_2$, chosen about the median of the extinction period. Any other sets of individual data might have been presented, but these give some view of the spread, as well as of the consistency of results of the various subjects.

In summary, the rate of stating opinions changed in accordance with the assumptions made. All *S*s increased their rate of stating opinions, regardless of the topic of conversation, its setting, or *S*'s particular relationship with the *E*. The order of magnitude of the effect depended upon the kind of reinforcement employed. How it may be related to the variables noted above cannot be inferred from the present data.

Discussion

Individual differences in the rates of speech, and of giving opinions, are most striking and highly significant. We have already noted that they are the joint outcome of *S*'s rate of speech, the length of his sentences, of *E*'s discrimination of his speech, and of *E*'s own speech rate. Of the two *S*s with the lowest rate of making statements, one was a Finn who spoke English with difficulty, and the other was a young woman who talked very fast and in very long sentences indeed. (She was also the most opinionated, according to our rate of giving opinions.) Since the experiment was performed, Fries's (1) work has become available, and a study of it suggests the basis of our *E*s' criteria.

The statements that the *E*s counted during the period of reinforcement are evidently identical with Fries's "utterance units" (1, p. 36), i.e., stretches of speech bounded by a change of speaker. During reinforcement and during extinction by disagreement, each stretch of *S*'s speech is bounded by *E*'s delivery of successive reinforcements or disagreements. The cues in *S*'s speech that determine *E*'s delivery of a reinforcement probably cannot yet be specified. However, the facts that the rate of uttering "statements" is stable, and that the rates reported by the same *E* are correlated with each other suggest that the "statements" or "sentences" counted during the operant level, and during extinction (although these are by definition not Fries's "utterances," since *E* says nothing) are stretches of speech such that *E* is *stimulated* to respond (1, p. 49). He does so, not by speaking, but rather by making a mark in his record. If this analysis is correct, then our *S*'s statements are what Fries also terms statements, i.e., "sentences that are regularly directed to eliciting attention to continuous discourse."

Magnitude of the Effect

These data do not permit us to draw conclusions about the magnitude of the effect, although it is clearly some function of the values of reinforcement variables. If *S* rarely states an opinion, it is difficult for the number of reinforcements to become very great, and the effect is necessarily small.

Acquisition Effects

The not-quite-significant difference in the median effects of paraphrasing and of simple agreement, and the significant difference in their variances are interesting. Probably many different kinds of paraphrases were employed; the differential effectiveness of these as reinforcing stimuli needs investigation. Both the smallest and the greatest changes in the rate of stating opinions were produced by paraphrasing.

Extinction Effects

During extinction by disagreement, some *S*s "marshalled the facts," others changed the topic. Some subjects who were extinguished by either treatment became "disturbed," or angry. There is more than a suggestion that when *S* undergoes complete nonreinforcement, his speech tends to extinguish and, indeed, he tends to leave the experimental situation earlier ("for study," "to go to dinner," and the like), but the 10-minute extinction period is too brief, and the variation among *E*s in continuing to record is too great to permit evaluation of this tendency.

General Remarks

Certain problems, soluble by further research, set limitations on the generality of the present results.

Only one of our *E*s was able to use a tape recorder, and clearly, the use of such an instrument, perhaps in conjunction with independent judges, might yield counts of all statements and opinion-statements that were less dependent on *E*'s own criteria. However, it is not at all clear that there would be less dependence on *E*'s criteria (1), since the delivery of reinforcements will necessarily continue to depend on *E*'s speech habits. A variety of specific utterances by *E* were employed as reinforcing stimuli; a study of the variability in the effectiveness of various kinds of statements by *E* would be most useful.

The present results do not permit us to state how important is the particular social relationship between *S* and *E*. Would agreement by an *E* whom *S* disliked reinforce his verbal behavior? These conversations were relatively short, with the result that extinction was carried out to its asymptote in only a few *S*s, and hence differences between the effect of disagreement and of complete nonreinforcement, although suggested, cannot be tested. Similarly, neither "satiation" effects of continuous and repeated reinforcement nor complete "talking-out" of *S* on a topic could occur. (It should be recalled that our procedure does not allow *E* to contribute anything new to the conversation.)

The topics of conversation were, in only a few cases, such that *S* might be

"ego-involved" in their outcome. Perhaps if *S* were subjected to these procedures when he was talking about something he "felt deeply" about, the results might differ, e.g., acquisition might be greater and extinction far slower. Orderly *changes* in the topic of conversation should also be observable (see footnote 5 on p. 104.)

Finally, it should be remembered that our *E*s were all well trained in conditioning before undertaking this experiment, and this experience may prove necessary for the successful completion of the experiment.

Despite these limitations, this experiment shows that if, in what is ostensibly an ordinary conversation, one agrees with opinions expressed by a speaker, the speaker will give still more opinions, and that returning the speaker's words in paraphrase has the same effect. It also shows that disagreement reduces the number of opinions given, as does ignoring the speaker's statement. The verbal behavior of a speaker, apparently without regard to its content or setting, is under the control, not only of the speaker himself, but also of the person with whom he is conversing.

These results are in accord with the two hypotheses made. But one may ask, is this operant conditioning? By any empirical, non-theoretical definition of conditioning, the changes in behavior found conform with those of conditioning, and the present results may be classified as conditioning. What are some of the alternatives?

Two can be noted, and both suggest that the data depend upon the *E*s' behavior, rather than the *S*s'. The *E*s may have "made up" the data, since they knew that certain kinds of data were expected of them. This alternative can be rejected without hesitation. The *E*s' previous performances, and the internal consistency of the data lend it no credulity. A second alternative is that "suggestion" may have altered the *E*s' discrimination of speech. If this were so, it would itself be a finding of interest. The writer is inclined to doubt very much that this occurred to any extent, in view of the phenomenon of "negative suggestibility," and of the frank skepticism of some *E*s as to the experiments' outcome before the data were collected and tabulated. Repetition of the experiment, with tape-recording of the verbal behavior of both *S* and *E* will permit ready evaluation of both these possibilities.

The results of this experiment make psychological and scientific sense of common-sense descriptions of conversation. ("People like to talk to people who are interested in what they are saying"; "if you ignore him, he'll go away"; "all right, if you don't believe me, here are the facts. . . .") and, indeed, other social and political behaviors. The data suggest that, once the appropriate simplifying assumptions are made, a very high degree of order can be revealed in "complex" situations, and that a still higher degree of order can be introduced into them.

The simplifying hypotheses made here are derived from the concepts of *response* and of *conditioning*, and they have proved experimentally fruitful in the present instance. This complex behavior is available to direct experimental

investigation, and the orderliness and lawfulness of the behavior exhibits itself when irrelevant details are ignored. The heuristic advantages of much of present stimulus-response theory, when it is applied in the field of verbal behavior in a social context, are clear.

If our interpretation is correct, experimental work on a wider variety of human social behavior is possible. The isolation in conversation of independent variables susceptible to direct manipulation and of dependent variables showing orderly change, should give a much wider and more significant scope to experimental investigation. The experiments now possible provide new techniques for the investigation of client-therapist relationships and of therapeutic techniques in clinical psychology. They may be applied to the study of the behavior of small groups, and of personality.

They suggest how cooperation may be ensured. They lead to questions such as, "Can one, by pairing oneself with a reinforcing stimulus, come to control effectively the behavior of a total stranger?" That is to say, if a person agrees with everything said by someone whom he has not previously known, will he then have other means of reinforcing, or of exerting other types of control over, the stranger's behavior? The possibilities are interesting.

Summary and Conclusions

Seventeen *E*s carried on conversations with 24 different *S*s.

Two assumptions are made, (*a*) that "stating an opinion" is a class of behavior that acts as a response, and (*b*) that statements of agreement with, or paraphrases of, such statements of a speaker act as reinforcing stimuli. From these it is inferred that the rate at which a speaker states opinions varies with the administration of agreement or of paraphrase by the person with whom he is conversing. The experimental conversations were carried out on a wide variety of topics of conversation, in a wide variety of places, and in a group of *S*s, most of whom were college students. The expected results appeared. Every *S* increased in his rate of speaking opinions with reinforcement by paraphrase or agreement. Twenty-one *S*s decreased in rate with nonreinforcement. Over-all rates of speaking did not change significantly.

In no case was the *S* aware that he was the subject of an experiment, or that the conversation was an unusual one.[5]

References

1. Fries, C. C. *The structure of English*. New York: Harcourt, Brace, 1952.
2. Skinner, B. F. The generic nature of the concepts of stimulus and response. *J. gen. Psychol.*, 1935, **12**, 40–65.
3. Verplanck, W. S. The operant conditioning of human motor behavior. *Psychol. Bull.*
4. Wilcoxon, F. *Some rapid approximate statistical procedures.* New York: American Cyanamid Co., 1949.

[5]The writer, after having described the experiment to someone in casual conversation, had the illuminating experience of then being used as *S* by the person to whom he had described it. He showed the effect and, like, it would seem, *all S*s in this experiment, was quite unaware that he had been an *S*.

Exercise

Improvisation and Attitude Change

A number of experimenters have reported that subjects who somehow become actively involved in a persuasive communication show more attitude change than other subjects who are merely passively exposed to the same communication. In one study (Janis and King, 1953), students who actually delivered a speech from an outline were more persuaded by their own speech than were members of their audience. In a second experiment (King and Janis, 1956) subjects who were required to improvise a speech in their own words showed more attitude change than subjects who read a similar prepared speech aloud. A third study (Janis and Mann, 1965) indicated that active role playing can affect behavior as well as attitudes. Subjects enacted a psychodrama in which they pretended that a doctor had just told them that they had lung cancer. These subjects who actively participated in the psychodrama reported that they smoked fewer cigarettes than comparison subjects who listened to a tape recording of the role-play session, but did not actively participate. Results from these and many other experiments indicate that at least in some situations speakers, debaters, and precinct workers convince themselves more than they persuade their intended target or audience.

This exercise provides four arguments for lowering the voting age to 17. Subjects in both the active participation and in the passive exposure condition will read the four arguments; in addition, subjects in the active participation group will be required to reformulate the arguments into an essay. On the basis of the research summarized above, you should expect the subjects who must improvise and reformulate the arguments to show more attitude change than those who merely read the original arguments on the same topic.

There are at least two separate theoretical reasons why this should occur. We can call the first reason *cognitive contact* (see, for example, Janis and Gilmore, 1964, for a current statement of that theory). Subjects who must actually use the information in order to write their own arguments may study it more carefully and think through the implications more thoroughly. They might spend more time rehearsing the arguments and learn them better, and they might be more motivated to discover additional arguments not contained in the initial message. In other words, they would have a greater cognitive contact with relevant information—especially arguments that are inconsistent with their own attitudes. Thus, cognitive contact with arguments supporting the "other side" could produce attitude change.

A second, quite different explanation follows from the theory of *cognitive dissonance* (Festinger, 1957; Aronson, 1969). If a subject thinks one thing privately, but knows that he is saying an opposing thing publicly, he holds two inconsistent cognitions. For example, he may believe that 17-year-olds should *not* be allowed to vote, but in an experiment or another social setting he may be induced to say that 17-year-olds *should* be allowed to vote. According to the theory of cognitive dissonance, the existence of these inconsistent, nonfitting, or dissonant cognitions is uncomfortable, and subjects will be motivated to reduce this uncomfortable dissonance. One way to reduce the dissonance between the cognitions "I *say* one thing," and "I *think* another" is to change one of the cognitions. A subject might change what he says to make it consistent with his attitudes, or he might change what he thinks to make it consistent with his behavior.

In this exercise, since you will induce some subjects to say what you want, it will be impossible for them to change their behavior; since those subjects will actually write essays counter to their initial attitudes, it will be difficult for them to deny or distort that fact to themselves. Thus, the easiest way to reduce the dissonance is to change their private attitude.

Your subjects *might* say something like this to themselves: "Gee, I feel kind of bad because I really don't think 17-year-olds should be allowed to vote, but I just wrote that essay that said they should. That really makes me uncomfortable. There certainly is no doubt about it that I wrote that essay saying that they should be allowed to vote. You know—now that I think about it—there really were some pretty good arguments in that essay I wrote. Now that I think about it, perhaps I'm not so sure that 17-year-olds shouldn't be allowed to vote."

Both cognitive contact and cognitive dissonance, then, should produce more attitude change for the subject who actually writes arguments compared to the subject who only reads the same arguments.

Procedure

First you will need to find subjects with a few minutes of time to spare. You might try laundromats, cafeterias, bus stations, or student lounges. If your instructor agrees, you can also approach people in dormitory rooms. When you locate a prospective subject, explain to him that your class is doing an attitude survey and ask him if he has just a few minutes to fill out a questionnaire. This is not a major imposition. If you, yourself, expect most people to participate, most of them will.

Three different experimental conditions will be described, but your instructor may prefer that you run only two conditions. It would be possible to run one of the experimental conditions and the control group or just the two experimental conditions.

(1) Comparison Condition. These subjects will do nothing except fill out the attitude questionnaire (p. 123). Record that a particular subject is in the control condition by circling the letter "C" on the questionnaire form. Since these subjects are not subjected to any experimental treatment (that is, they will not read or write a persuasive communication) their attitude scores should reflect the opinion of the "average man" in your sample. We can use these scores to estimate how the experimental subjects would have responded if you had given them a test *before* (a pretest) they were subjected to a persuasion attempt (see Introduction, p. 19 regarding posttest-only designs). Record the subject's questionnaire responses on Data Sheet 1, columns 2 and 3.

(2) Read Outline Condition. Following the script word for word (p. 113), you will tell the subject, after he has agreed to take the opinion poll, that the class has prepared some fact sheets on the issue of whether to allow 17-year-olds to vote. You will give him the outline which is included on page 115. As you can see, page 115 includes some arguments in favor of lowering the voting age and some questions for the subject to answer regarding the "fact sheet." After explaining that the subject's own attitudes might affect his ratings on the "fact sheet," administer the questionnaire which includes two questions on the 17-year-old vote issue. If subjects in this *read outline* condition are more in favor of 17-year-old voting than control subjects, then reading the outline must have made them more favorable. Both groups were treated exactly alike except for the fact that the *read outline* experimental group read the "fact sheet" and the *control* group did not. So if the attitudes of the two groups are different statistically (more different than might reasonably be expected by chance), then it must have been reading the outline

which caused the difference. Record responses on Data Sheet 2, columns 13 and 14.

(3) Read and Write Condition. Again follow the script (p. 113) and explain that the class is collecting arguments in favor and arguments against the 17-year-old vote. Continue that the class has discovered that the best way to get arguments is to have people write on only one side of the issue, so you are asking some people to write in favor and other people to write against the issue. You will explain that the subject can choose the side he wants, but you already have enough arguments against the 17-year-old vote, and it would be better if the subject wrote a very short essay in favor of lowering the voting age to 17.

When the subject agrees, give him the fact sheet on page 119. It contains the same fact sheet as you will give the *read outline* subjects, but it also contains a space for the subjects to add improvised arguments. At the bottom of the page is a space for subjects to rate their own essay along the same dimensions the *read outline* subjects used to rate their arguments. Then administer the questionnaire in the same manner as for the *read outline* subjects. Record responses on Data Sheet 3, columns 13 and 14.

If the attitudes of this group are significantly more favorable than in the control group, then reading, writing, or a combination of the two, produced the attitude change. If these attitudes are more favorable than those in the *read outline* condition, the writing, per se, must have produced more attitude change than reading only.

Data Analysis

Strictly speaking the data from the present experiment should be analyzed by statistical techniques called analysis of variance and Dunnet's t (Winer, 1962); but the simpler t-test procedures described here are a reasonable approximation. Copy the answers from the questionnaires on to Data Sheets 1, 2, and 3.

(1) Test the read-only experimental group against the comparison group with a t test (Appendix A). This will tell you if reading the arguments changed attitudes. The mean, sum X and sum X^2 terms can be computed on the data form for the t-tests on Data Sheets 1, 2, and 3.

(2) Test the read and write condition against the comparison group with a t-test (Appendix A). This will tell you if the combination of reading the arguments and improvising the essay produced any attitude change.

(3) Test the read-only experimental against the read and write experimental group with a t-test (Appendix A). This will tell you if the improvisation had an effect over and beyond the effect produced by just reading the arguments.

Optional Further Analyses

(4) Compute the rank order correlation (Appendix C) between rating of arguments (column 9 of Data Sheet 2) and final attitude (column 4) for the read condition. This will tell you if subjects who liked the arguments also wanted the voting age lowered to 17.

(5) Compute the rank-order correlation (Appendix C) between final attitude (column 15) and ratings of own essays (column 20). This will tell you if people who thought they wrote good essays are more in favor of lowering the voting age to 17.

(6) What are the correlations between the two attitude questions (Q_1 and Q_2) which were added together to form a single index number? (*S*s from all conditions can be grouped together for this analysis.)

(7) All of the above analyses on attitude have been done with a summary index created by adding questions 1 and 2. All the above analyses can be repeated for each attitude item separately. In other words, you can do a *t*-test using only Q_1 or only Q_2 as a measure of attitude.

References

Hovland, C., Janis, I., and Kelly, H. H. *Communication and persuasion.* New Haven, Yale University Press, 1953.

This book contains a chapter on active participation which describes the pioneering Yale studies. In one experiment (King and Janis, 1953), subjects who actually deliver a speech are more persuaded than members of their audience. In the second experiment, which is included as the Research Example for this Exercise, (Janis and King, 1956), subjects who must improvise from notes are more persuaded by their presentation than subjects who merely read a prepared speech aloud.

Janis, I., and Mann, L. "Effectiveness of emotional role-playing in modifying smoking habits and attitudes." *Journal of Experimental Research in Personality,* 1965, 1, 84–90; Mann, L. and Janis, I. "A follow-up study on the long-term effects of emotional role playing." *Journal of Personality and Social Psychology,* 1968, 8, 339–342.

These two experiments describe role-playing procedure which is effective in persuading subjects to decrease the amount they smoke. The effect is shown to last over a long period of time.

Carlsmith, J. M., Collins, B. E., and Helmreich, R. L. "Studies in forced compliance: I. The effect of pressure for compliance on attitude change produced by face-to-face role playing and anonymous essay writing." *Journal of Personality and Social Psychology*, 1966, 4, 1–13.

Recently, a theoretical controversy has arisen concerning the effects of inducements, usually money, used to entice subjects to engage in the conterattitudinal role playing. (Counterattitudinal role playing occurs when a subject is asked to take a role which is inconsistent with his own attitudes.) Some experiments have found that giving students a large incentive, e.g., $20.00 to engage in role playing facilitates the amount of attitude change, but other experiments have found exactly the opposite effect. But the major concern of this exercise is with the basic phenomenon that active participants show more attitude change than subjects passively exposed to information. Carlsmith, Collins, and Helmreich is one of the many studies on magnitude of inducement and forced compliance. In one of their experimental conditions, a large inducement to engage in role playing facilitates attitude change and in another condition it inhibits attitude change. The paper reviews the theoretical issues involved in this controversy as well as previous research.

Elms, A. *Role playing, reward, and attitude change.* New York: Van Nostrand Reinhold, 1969. A paperback collection of articles on improvisation including almost all of those cited in the present exercise.

Janis, I. and Gilmore, B. The influence of incentive conditions on the success of role-playing in modifying attitudes. *Journal of Personality and Social Psychology*, 1965, **1**, 17–27.

The introductory section of this article contains a good statement of cognitive contact theory.

Festinger, L. *A theory of cognitive dissonance.* Evanston, Ill.: Row, Peterson, 1957. This book contains the original statement of dissonance theory.

Aronson, E. Dissonance theory: Progress and problems. In R. P. Abelson, E. Aronson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, and P. Tannenbaum : (Eds.), *Theories of cognitive consistency: A sourcebook.* Chicago: Rand McNally, 1968, p. 5–27.

This article is a more recent attempt to update dissonance theory.

Experimenter's Summary Sheet and Working Script

Take this sheet along with you when you run your subjects. You should memorize the script, but you may keep it in front of you on a clipboard to refresh your memory.

Write the initials of each experimental condition (C, R, and RW) on three pieces of paper. Determine a random order to run the experimental conditions by blindly drawing the three sheets of paper out of a container. Then replace the papers and draw again to determine a new random order for the next set of three conditions, etc. Make a list of the order of conditions for all the subjects you intend to run before you begin.

READ OUTLINE CONDITION

"Hello, my name is ________________________. My social psychology class is conducting a survey of attitudes on several issues. Would you be willing to take a few minutes to fill out a short questionnaire?

"Our class has prepared several fact sheets which we plan to use in an experiment later. The topic is whether to lower the voting age to 17. The first thing I would like you to do is read this fact sheet and answer the questions about it at the end of the page." (When he has finished, hand him the questionnaire.) "As you might imagine, your own attitude might affect how you rate the 'fact sheet'; so we would also like you to fill out this questionnaire on your own attitudes." (After questionnaire is completed, circle the "R" on the form.)

SCRIPT FOR READ AND WRITE CONDITION

"Hello, my name is ________________________. My social psychology class is conducting a survey of attitudes on several issues. Would you be willing to take a few minutes to fill out a short questionnaire?

"Our class is preparing several fact sheets which we plan to use in an experiment later. The topic is whether or not to lower the voting age to 17. We will use the arguments in a persuasive communication which the class is writing. We have found that the best way to get arguments on both sides is for some people to write on one side and for other people to write on the other side. *The choice is completely up to you*, but I already have enough arguments against the 17-year-old vote. Would you be willing to write a short essay in favor of lowering the voting age to 17?

"Here is a sheet for your essay. First look at the arguments at the top of the page, to get a few ideas for your own essay." (Hand subject p. 119.)

"After you have written a brief essay, answer the questions at the bottom of the page."

(When he has finished, hand him the questionnaire.) "As you might imagine, your own attitudes might affect how you rate your own essay and the kind of essay you write; so we would also like you to fill out this questionnaire on your own attitudes." (After questionnaire is completed, circle the "RW" on the questionnaire.)

Form R Name ______________________________

Arguments in Favor of Lowering the Voting Age to 17

1. If you are old enough to fight, you are old enough to vote.

2. Young people would be less likely to riot and cause disturbances on the campus if they could express their feelings in the ballot box.

3. Young people would be more responsible in general if they knew they had the responsibility to decide how to vote.

4. Our improved public education means that nowadays the average 17-year-old is much better informed than they were when the voting age was set at 21.

I think the above arguments are (circle one number):

1	2	3	4	5	6	7
Extremely Persuasive	Somewhat Persuasive	Moderately Persuasive	Neither Persuasive nor Unpersuasive	Moderately Unpersuasive	Somewhat Unpersuasive	Extremely Unpersuasive

I think the above arguments are (circle one number):

1	2	3	4	5	6	7
Extremely Logical	Somewhat Logical	Moderately Logical	Neither Logical nor Unlogical	Moderately Unlogical	Somewhat Unlogical	Extremely Unlogical

Form R Name ______________________________

Arguments in Favor of Lowering the Voting Age to 17

1. If you are old enough to fight, you are old enough to vote.
2. Young people would be less likely to riot and cause disturbances on the campus if they could express their feelings in the ballot box.
3. Young people would be more responsible in general if they knew they had the responsibility to decide how to vote.
4. Our improved public education means that nowadays the average 17-year-old is much better informed than they were when the voting age was set at 21.

I think the above arguments are (circle one number):

1	2	3	4	5	6	7
Extremely Persuasive	Somewhat Persuasive	Moderately Persuasive	Neither Persuasive nor Unpersuasive	Moderately Unpersuasive	Somewhat Unpersuasive	Extremely Unpersuasive

I think the above arguments are (circle one number):

1	2	3	4	5	6	7
Extremely Logical	Somewhat Logical	Moderately Logical	Neither Logical nor Unlogical	Moderately Unlogical	Somewhat Unlogical	Extremely Unlogical

Form RW Name________________________________

Arguments in Favor of Lowering the Voting Age to 17

1. If you are old enough to fight, you are old enough to vote.
2. Young people would be less likely to riot and cause disturbances on the campus if they could express their feelings in the ballot box.
3. Young people would be more responsible in general if they knew they had the responsibility to decide how to vote.
4. Our improved public education means that nowadays the average 17-year-old is much better informed than they were when the voting age was set at 21.

Short essay in favor of lowering the voting age to 17:

1. I think my essay is (circle one number):

1	2	3	4	5	6	7
Extremely Persuasive	Somewhat Persuasive	Moderately Persuasive	Neither Persuasive nor Unpersuasive	Moderately Unpersuasive	Somewhat Unpersuasive	Extremely Unpersuasive

2. I think my essay is (circle one number):

1	2	3	4	5	6	7
Extremely Logical	Somewhat Logical	Moderately Logical	Neither Logical nor Unlogical	Moderately Unlogical	Somewhat Unlogical	Extremely Unlogical

Form RW Name ____________________

Arguments in Favor of Lowering the Voting Age to 17

1. If you are old enough to fight, you are old enough to vote.
2. Young people would be less likely to riot and cause disturbances on the campus if they could express their feelings in the ballot box.
3. Young people would be more responsible in general if they knew they had the responsibility to decide how to vote.
4. Our improved public education means that nowadays the average 17-year-old is much better informed than they were when the voting age was set at 21.

Short essay in favor of lowering the voting age to 17:

1. I think my essay is (circle one number):

1	2	3	4	5	6	7
Extremely Persuasive	Somewhat Persuasive	Moderately Persuasive	Neither Persuasive nor Unpersuasive	Moderately Unpersuasive	Somewhat Unpersuasive	Extremely Unpersuasive

2. I think my essay is (circle one number):

1	2	3	4	5	6	7
Extremely Logical	Somewhat Logical	Moderately Logical	Neither Logical nor Unlogical	Moderately Unlogical	Somewhat Unlogical	Extremely Unlogical

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.

Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.

Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.
Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.
Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.
Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Opinion Questionnaire

Interviewer, fill in the following:

Interviewer number ______
Circle one: C, R, RW.
Please circle the number that most closely represents your opinion.

Seventeen-year-olds should be able to vote:

1	2	3	4	5	6	7
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Individuals must be at least 21 before they are mature enough to vote.

7	6	5	4	3	2	1
Agree Definitely!	Agree	Agree Slightly	Neither Agree nor Disagree	Disagree Slightly	Disagree	Disagree Definitely!

Data Sheet 1
Control Condition

Subject Number	Q_1	Q_2	$Q_1 + Q_2$	$(Q_1 + Q_2)^2$
1	2	3	4	5

Add all numbers in column and enter in boxes

Sum X	Sum X^2

$$\text{Mean} = \frac{\text{Sum } X}{N}$$

Data Sheet 2
Read Only Condition

Subject Number	Opinion Questions					Ratings of Arguments				
	Q_1	Q_2	$Q_1 + Q_2$	$(Q_1 + Q_2)^2$	Rank of $(Q_1 + Q_2)^2$	Q_1	Q_2	$Q_1 + Q_2$	$(Q_1 + Q_2)^2$	Rank of $(Q_1 + Q_2)^2$
1	2	3	4	5	6	7	8	9	10	11

Add all numbers in column and enter in boxes

Sum X Sum X^2 Sum X Sum X^2

$$\text{Attitude Mean} = \frac{\text{Sum } X}{N}$$

$$\text{Rating Mean} = \frac{\text{Sum } X}{N}$$

Data Sheet 3
Read and Write Condition

Subject Number	Opinion Questions					Ratings of Arguments				
	Q_1	Q_2	$Q_1 + Q_2$	$(Q_1 + Q_2)^2$	Rank of $(Q_1 + Q_2)^2$	Q_1	Q_2	$Q_1 + Q_2$	$(Q_1 + Q_2)^2$	Rank of $(Q_1 + Q_2)^2$
12	13	14	15	16	17	18	19	20	21	22

Add all numbers in column and enter in boxes

	Sum X	Sum X^2		Sum X	Sum X^2

Attitude Mean $= \frac{\text{Sum } X}{N}$ ☐

Rating Mean $= \frac{\text{Sum } X}{N}$ ☐

Research Example

Comparison of the Effectiveness of Improvised Versus Non-Improvised Role-Playing in Producing Opinion Changes[1]

BERT T. KING AND IRVING L. JANIS

The judgments, opinions, and attitudes that one overtly expresses to friends and associates frequently come from norm-setting communications to which all persons who have a given status position within the community are expected to conform. In order to live up to social role expectations, a person will often repeat a message to others as if it were his own position, even though it may not be in accord with his private convictions. Experimental evidence, reported by the authors in an earlier paper (3), indicates that role-playing activity can exert a marked influence on the individual's private opinions. Further evidence bearing on the way in which outer conformity affects inner conformity will be presented in the present report, which describes an experiment designed to test two explanatory hypotheses suggested by the earlier findings.

In the earlier experiment we obtained measures of the effectiveness of three persuasive communications, each of which was presented under conditions of active role-playing and passive exposure. A greater amount of opinion change was observed when subjects were induced to play the role of a communicator who is attempting to convince others of the arguments and conclusions, than when

SOURCE. *Human Relations* 1956, **9**, 177–186.

[1] This study was conducted at Yale University as part of a program of research on factors influencing changes in attitude and opinion. The research program was supported by a grant from the Rockefeller Foundation under the general direction of Professor Carl I. Hovland, to whom the authors wish to express appreciation for helpful suggestions and criticisms.

they were allowed to remain passive members of an audience who merely read and listened to the material. Thus, the evidence indicates that overt verbalization induced through role-playing tends to augment the effectiveness of a persuasive communication. This outcome seems to bear out the notion that "saying is believing".

The psychological processes activated when a person is participating in the role of a communicator may involve many different factors that could affect his degree of attention, comprehension, and motivation to accept the content of the persuasive message. From among the numerous factors that theoretically could contribute to the superiority of active participation over passive exposure, we have singled out two for systematic investigation—"improvisation" and "satisfaction". In our earlier experiment, there were some supplementary observations and suggestive correlational data that led us to formulate an "improvisation" hypothesis and a "satisfaction" hypothesis as alternative explanatory concepts to help account for the observed gain from role-playing.

The "improvisation" hypothesis asserts that people tend to be especially impressed by their own persuasive efforts when they are stimulated to think up new arguments and appeals in order to do a good job of convincing others. According to this hypothesis, a person will end up by convincing himself of the validity of the point of view he is required to defend, provided that he has been induced to improvise new supporting ideas.

The "satisfaction" hypothesis postulates that the individual's sense of achievement or satisfaction with his performance in a social role provides a special source of reward that is capable of reinforcing the opinions he overtly verbalizes in that role. According to this hypothesis, role-playing will have a positive effect with respect to changing private opinions in circumstances in which the individual feels satisfied with his overt performance.

In order to investigate the relative importance of the "improvisation" and the "satisfaction" factors, the present experiment was set up to compare the effects of role-playing under laboratory conditions such that each of the two factors would be varied through direct experimental manipulation.

Methods and Procedures

The subjects (male undergraduate students) were assigned at random to various experimental groups, each of which was exposed to the same persuasive communication. Written in the style of ordinary mass-media magazine articles, the communication presented arguments in support of two main conclusions: (*a*) that over 90 per cent of college students will be drafted within one year of their graduation, and (*b*) that the majority of college students will be required to serve at least three years in the armed forces (i.e. one year longer than the present requirement). Since all subjects were eligible for military service, it was assumed that this topic, unlike the relatively impersonal ones used in the earlier experiment, would be of considerable personal concern to them.

Opinion changes were observed by obtaining each subject's opinion re-

sponses (on a series of five questions) immediately after the communication. These answers were compared with the answers the subject had given to the same question in an opinion survey that had been conducted several months earlier.

The role-playing instructions were similar to those used in our earlier experiment (3). The procedure was modelled after those described by Lippitt (5), Maier (6), Moreno (7), and other investigators who have employed psychodramatic situations requiring the subject to enact a role in which he expresses beliefs, judgments, and attitudes that are not necessarily compatible with his own private convictions. In the present experiment, the subjects were requested to play the role of a sincere advocate of the point of view expressed in the communication on military service, which they read silently beforehand. Each active participant was told that the purpose of this role-playing activity was to help us develop a new oral speaking test and that his talk would be tape-recorded for later presentation to a group of judges. Each passive control subject, on the other hand, merely read the script silently.[2]

In order to assess the effects of the improvisation and satisfaction factors, two different groups of active participants were used: Group A was given a relatively difficult task that required a high degree of improvisation. The subjects in this group were required to play the role of an impromptu speaker, presenting the talk without the written script, immediately after having read it silently. Group B was assigned a much easier role, which could be expected to heighten the degree of satisfaction—that of a speaker who reads aloud from the completely prepared script without being required to do any improvising. In effect, comparison of the two groups pits the improvisation factor (represented by Group A) against the satisfaction factor (represented by Group B).

Results

In order to determine whether the experimental treatments succeeded in eliciting different degrees of satisfaction, every subject in Groups A and B was asked to rate four different aspects of his own oral speaking performance immediately after he had given the talk. The results presented in Table 1 indicate that

[2] The level of attention and related factors that might be a function of the "set" to deliver an oral presentation were controlled for both groups through the supplementary use of an irrelevant communication. Each of the active and passive subjects was asked to read to himself two different scripts, and was told that when finished he would be required to give a talk based on one or the other of the two communications. After this standard preparation, the active participants were requested to give a full-length talk based on the communication that discussed the prospects of military service, while the passive controls were asked to deliver a brief talk based on the other (irrelevant) communication. Thus the procedure was designed to elicit a comparable set from the passive controls and the active participants during the primary exposure to the main communication. Moreover, a test covering the essential informational content of the communication was included along with various opinion measures in the post-communication questionnaire, so as to determine whether the additional rehearsal involved in giving the talk (cf. 2, pp. 263-5) had any observable effect on learning.

Table 1

Percentage of the Improvisation and Non-Improvisation Groups Giving Favorable Self-Ratings on Oral Speaking-Performance

Self-rating	*Group A (Improvisation) (N = 32)*	*Group B (Non-improvisation) (N = 23)*	*% Diff. (B − A)*	*p-value*
	%			
1. Participant felt his performance was at least "satisfactory"	63	96	33	$<.01$[a]
2. Participant felt he did *not* distort any arguments	53	96	43	$<.01$
3. Participant felt his voice was *not* monotonous	65	91	26	.01
4. Participant felt he gave impression of being "sincere"	47	70	23	.04

[a]Notations such as $p < .01$ are explained in the Introduction, pp. 15–16.

on each of the four items the percentage giving favorable self-ratings was significantly higher in Group B than in Group A. This evidence verifies our assumption that the more difficult task of improvising assigned to Group A would make for more self-criticism and dissatisfaction with respect to the individual's oral speaking-performance. Hence, the prediction from the "satisfaction" hypothesis is that Group B should show significantly more opinion change than Group A.

The "improvisation" hypothesis makes the opposite prediction. Among the subjects in Group A, all of whom were required to formulate the communication in their own words, a substantial percentage introduced new arguments or original elaborations of the prepared script into their talks, whereas the nature of the task assigned to Group B precluded such improvisation. If the improvisation factor is a critical one in mediating the effects of role-playing, Group A should be found to show significantly more opinion change than Group B.

The opinion change data for the two groups of active participants and for the group of passive controls are presented in Table 2. The results, based on the five key questions included in the pre- and post-communication questionnaires, show that Group A was more influenced by the communication than was Group B. This trend is apparent on all five opinion items. The superiority of Group A is reliable at the .01 confidence level on the combined index of change, which shows the percentage of those in each group who changed on three or more items in the direction of the conclusions advocated by the communication. Moreover, Group A showed more opinion change than Group C (the passive control group) on four of the five items and differed reliably from Group C on the combined index. Thus, the results sup-

Table 2
Opinion Changes Produced Under Improvised and Non-Improvised Role-Playing Conditions

Opinion Items[a]	Net opinion change: percentage changing in the direction advocated by the communication minus percentage changing in opposite direction		
	Group A (Improvisation) (N = 32)	*Group B (Non-improvisation) (N = 23)*	*Group C (Passive Exposure) (N = 20)*
	%	%	%
1. Estimates of required length of service for draftees	41	27	5
2. Estimates of college students' chances of being deferred	44	26	25
3. Estimates of college students' chances of becoming officers	70	47	45
4. Expectations concerning the length of one's own military service if drafted	59	46	50
5. Expectations concerning the length of one's own deferment before being drafted	50	26	55
Combined index: percentage influenced on three or more of the five opinion items	87½	54½	65
	A vs. B: p-value = .01		
	A vs. C: p-value = .03		

[a] Copies of the questionnaire (showing the exact wording of each of the five key questions) are available upon request from the Institute of Human Relations, Yale University.

port the "improvisation" hypothesis, since they indicate that there is a significant gain from active participation when the individual is required to engage in improvised role-playing.

Predictions from the "satisfaction" hypothesis, on the other hand, are not confirmed. First of all, Group B failed to show as much opinion change as Group A, despite the fact that the subjects in the former group expressed a markedly higher degree of satisfaction. Secondly, Group B did not differ significantly from the passive control group on any of the five items; on the combined index there is a non-significant difference in the reverse direction ($p > .30$). Thus, there was no observable gain in opinion change from the relatively satisfying form of active participation in which Group B engaged.

A third set of pertinent data was ob-

tained from a supplementary procedure that had been introduced into the experiment to provide an independent measure of the effects of the "satisfaction" variable. Within the improvisation group, different degrees of satisfaction were elicited by means of performance ratings that were given by the experimenter at the end of each subject's talk. On a random basis, Group A was divided into three sub-groups, one receiving favorable ratings, another unfavorable ratings, and the third no ratings. In effect, the experimenter differentially administered social rewards and punishments for the purpose of varying the satisfaction variable under conditions in which improvisation was held constant.

As expected, self-ratings were found to be significantly affected by the experimenter's ratings. But on the five key opinion items the subjects who received favorable ratings (and expressed the highest degree of subjective satisfaction) showed approximately the same amount of opinion change as those who received unfavorable ratings or no ratings. These supplementary findings together with the data in Table 2 indicate that there was no observable gain in opinion change when feelings of satisfaction were experimentally induced by two different methods.

Discussion

The results from this experiment substantiate the observations on the importance of improvisation noted in our earlier experiment, but fail to confirm the findings on the relationship between degree of satisfaction and amount of opinion change. Since in the present study the experimental conditions were actually manipulated so as to induce variations in the degree of satisfaction, more confidence may be placed in these results than in the correlational evidence from the first experiment. The weight of the evidence now available clearly favors the improvisation variable as being a more important determinant of role-playing effects. Mere repetition of a persuasive communication, even under very favorable circumstances, apparently has little or no effect as compared with an improvised restatement of the message.

This tentative conclusion still leaves open the theoretical question as to how and in what particular way improvisation heightens the acceptance of new ideas. One possibility is that when a person engages in improvised role-playing his learning efficiency is improved because of increased attention. Data bearing on the effects of variations in the level of attention, discussed in our earlier report (3), suggest that this variable probably is not a crucial factor that could account for the augmented opinion change produced by experimentally induced role-playing. In the present experiment, additional findings pertinent to the attention variable were obtained from the recall test, which had been given to all our subjects shortly after they had been exposed to the main communication. If level of attention is a determining factor, the active participants, and particularly the improvisers, should have benefited from heightened learning efficiency and therefore should have obtained higher recall scores than the passive controls. But the results show that the

improvisation group, as well as the oral reading group, did *not* obtain higher recall scores than the group that was passively exposed to the communication. Hence, it seems unlikely that attention factors—or any related factors that operate through raising the level of learning efficiency—could account for the observed gain in opinion change produced by improvised role-playing.

Another type of explanation is suggested by the fact that improvisation requires the participant to reformulate the communication in his own words. The mere act of translating the message into one's own more familiar vocabulary might make the communication more meaningful, perhaps by ensuring that the implications of the content will be better understood and more easily assimilated into one's pre-existing framework of beliefs, attitudes, and values. But, to the extent that the reformulation factor is assumed to operate by facilitating the audience's comprehension of the communication, it does not seem to offer a plausible explanation of the outcome of our two role-playing experiments. The communications used in the first experiment and in the present one generally relied on relatively simple and familiar arguments that presumably could be easily understood by the college students who served as subjects. In the earlier experiments, one of the three communications contained a number of complicated arguments that made use of technical scientific concepts; but this communication proved to be the one with respect to which active participation had the least effect.

Since neither the attention hypothesis nor the reformulation hypothesis appears to be a satisfactory explanation, we must seek for some other characteristic of improvisation that may provide a more promising lead. For the present, it seems likely that the critical variable has to do with the *inventive* aspect of improvising. Here we are referring to one of the most salient features of an improvised performance: the spontaneous additions and elaborations of the arguments contained in the communication. This characteristic was found to be associated with the amount of opinion change in both of our role-playing experiments and in Kelman's experiment on the effects of a persuasive communication on school children who were induced to write essays in conformity with the communicator's position (4).

It seems plausible that there is a lowering of psychological resistance whenever a person regards the persuasive arguments emanating from others as his "own" ideas. This assumption might help to explain the effects of improvised role-playing. Hollingworth (I) contends that the effectiveness of a suggestion depends in part on the extent to which it appears to be of personal origin: resentment and negativistic reactions may interfere with acceptance of a *direct* suggestion from others, whereas the individual's belief that he is making a decision on his own initiative may increase the influence of an *indirect* suggestion. The notion that a direct approach tends to stimulate internal resistance seems to be a major assumption in theoretical discussions of the rationale for nondirective psychotherapy (8).

In the present experiment, impro-

vised role-playing might have been successful in helping to overcome resistance by reducing the intensity of those internal responses which normally interfere with the acceptance of persuasive messages. Among the major types of interfering response that one might expect to find in everyday communication situations would be doubts about the communicator's trustworthiness, thoughts about opposing arguments, and conflicting anticipations concerning the consequences of adopting the communicator's position. Any device that successfully decreases the occurrence of such interfering responses could be expected to heighten acceptance of a persuasive communication. Perhaps it is in this way that improvised role-playing facilitates opinion change.

When passively exposed to a persuasive communication, many persons may fail to be convinced because, although capable of fully comprehending the meaning of the arguments, they fail to have the sort of thoughts or anticipations that would motivate them to change their minds. Consider, for example, the communication on the prospects of military service for college students used in the present experiment. One of the main arguments was that college students are urgently needed in the military service because of a critical shortage of skilled personnel. When someone merely reads this argument he may think of it in purely abstract terms, wonder whether it is really true, and remain unconvinced. But if the same person is required to play a role in which he must "put this idea across" to others, he may be less likely to think of criticisms or objections and more likely to experience the convincing thought-sequences and vivid anticipations that will incline him to accept a new position on the issue.

Improvised role-playing might be viewed as a technique that induces the recipient to contribute to making the communication as effective as possible; he is stimulated to think up new arguments, cogent illustrations, and impressive appeals that will help to "sell" the conclusion. In effect, the customer is not simply asked to examine the ready-made material in the original communication but is given scissors, needle, and thread to hand-tailor the material to suit himself.

Summary and Conclusions

This report deals with an experimental investigation of the conditions under which inner beliefs or opinions are affected when one is induced to become an active participant with respect to communicating a persuasive message. The hypotheses tested were derived from an earlier experiment (3), which showed that subjects who were required to play a role in which they verbalized a communication aloud to others tended to be more influenced than those who were passively exposed to the same communication. The present experiment was designed to assess the importance of two factors that could mediate this role-playing effect: (*a*) improvisation of one's own arguments in support of the assigned conclusion; and (*b*) satisfaction with one's own speaking-performance.

Three equivalent groups of college students were given the same persuasive communication, which took the position that they would soon be drafted into the armed forces and would be required to serve a year longer than current draftees. The passive controls merely read the communication silently to themselves. One group of active participants read the script aloud, and a second group was required to give an improvised talk after having silently read the script. The oral reading-task, which involved no improvisation, evoked a markedly higher degree of satisfaction than the more difficult improvisation task, as indicated by the students' self-ratings. Additional variations in degree of satisfaction were introduced by subdividing the improvisation group so that some received favorable ratings from the experimenter on their speaking-performance while others received unfavorable ratings or no ratings.

The improvisation condition was found to be the only experimental variation that produced a significant increase in personal acceptance of the persuasive communication. The results consistently indicate that the amount of opinion change produced through active participation is dependent upon the amount of improvisation, but is not related to amount of satisfaction. Various psychological mechanisms were discussed that might help to explain the importance of improvisation in transforming outer conformity into inner conformity.

References

1. Hollingworth, H. L. *The Psychology of the Audience.* New York: American Book Co., 1935.
2. Hovland, C. I., Lumsdaine, A. A., and Sheffield, F. D. *Experiments on Mass Communication.* Princeton: Princeton University Press, 1949.
3. Janis, I. L., and King, B. T. "The influence of role-playing on opinion-change." *J. abnorm. soc. Psychol.*, 1954, Vol. **49**, pp. 211–18.
4. Kelman, H. C. "Attitude change as a function of response restriction." *Hum. Relat.*, 1953, Vol. **6**, pp. 185–214.
5. Lippitt, R. "The psychodrama in leadership training." *Sociometry*, 1943, Vol. **6** pp. 286–92.
6. Maier, N. R. F. *Principles of Human Relations*. New York: Wiley, 1952.
7. Moreno, J. L. *Psychodrama*. Vol. I. New York: Beacon House, 1946.
8. Rogers, C. R. *Counseling and Psychotherapy*. New York: Houghton Mifflin, 1942.

Exercise

Similarity, Attraction, and Social Influence

Similarity, attraction, and social influence are closely related variables. Numerous experiments dealing with these three variables indicate that an increase in any one of them tends to produce an increase in the other two. Thus, if two people become more similar (or if they *think* they have become more similar), they will generally like each other more and also have more influence on each other (Newcomb, 1960, 1961). Conversely, if two people increase their influence on one another, they will become more similar and usually like each other more; and if two people come to like each other more, they will become more similar and have more influence over one another.

A good example of research in this area is a study done by Sampson and Insko (1964) in which they experimentally demonstrated the relationship between attraction and influence. They trained a confederate to behave either very nicely or very offensively with a subject. As expected, the confederate's behavior seemed to induce subjects to like him when he was agreeable and cooperative and to dislike him when he was offensive. When the confederate later attempted to influence subjects, he was more successful when he was liked than when he was disliked.

In another experiment, Back (1951) employed three different types of instructions to create liking between partners working on a two-man task. The instructions emphasized either (1) the potential attractiveness of the partner, (2) the importance of the shared task, or (3) the prestige of the two-man unit. First each subject was asked to write a story about three pictures, then to discuss the pictures with his partner, and finally to write a second version of the story. A subject was rated as influential to the extent

that his ideas were adopted in his partner's final story. Each of the three sets of instructions were shown by the results to facilitate both liking and influence between partners.

The following exercise studies the same three variables as the preceding experiments—that is, similarity, attraction, and influence. The purpose is to test the hypothesis that a communicator is more influential (as measured by attitude change) when the receiver *likes* the communicator and/or when the receiver sees himself as being *similar* to the communicator.

Procedure

Recruit potential subjects in laundromats, bus stations, dormitories, or the library. You will probably need at least 8 to 10 subjects in each condition. (As always, more subjects increase the probability of obtaining statistical significance.) Tell them that you are conducting a 15-minute experiment on the effectiveness of interpersonal communication. After the subject has agreed to participate, you will give him the Values Questionnaire (p. 153), which includes several general questions. After pretending to score this questionnaire, tell the subject that he is either very dissimilar or very similar to the author of the communication that you are using. You will then give the subject a brief persuasive communication (p. 150) ostensibly written by someone who was contacted in the same manner in which the subject was contacted (for example, by another experimenter in a laundromat if you are running subjects in laundromats). After the subject has read the persuasive communication, he will be asked (1) to rate the author and (2) to indicate his own attitudes of the topic in the persuasive message (Questionnaires p. 155 and 156). The Experimenter Summary Sheet (p. 151-152) includes verbatim scripts for recruiting and instructing subjects. Before continuing to read, study the suggested script and procedure for the two experimental conditions and the comparison condition.

Deciding About the Manipulation

Ideally, two experimental groups should differ in one and only one respect. If the experimental manipulation of similarity produces different amounts of attitude change between the two groups, we would then know that perceived similarity was the factor that led to the differential attitude change. If no differences between groups were noted, and assuming that the experiment was well conceived and executed, we could conclude that communicator similarity had little or no effect on attitude change. However, when experimental results are negative (that is, when there is no difference between groups) we must consider the possibility that the manipulation of the independent variable (in this case, similarity) was too subtle or too weak

to produce measurable effects on the dependent variable (attitude change). It is quite possible that the one sentence or phrase suggesting that the author of the communication is similar or dissimilar to the subject will go unnoticed and thus produce no effects on the subject's attitudes.

A common solution to this problem is to strengthen the manipulation by increasing the differences between the two experimental groups. Your class, for instance, might decide to strengthen the similarity manipulation by adding the phrase, "I'm sure that you would like this fellow very much because he's your kind of person," to the "similar" condition and the phrase, "I'm sure you would not like this fellow because he's not at all your kind of person," to the "dissimilar" condition. By magnifying the differences between the two experimental conditions, you may be more likely to produce different amounts of attitude change.

However, when you increase the *impact* of your experimental manipulation by adding material, you may lose *clarity* in the interpretation of your results, since you become less certain of exactly what caused the observed effect. If you do obtain differences between groups of subjects, ask yourself the following questions: Are the differences because of the similar-dissimilar manipulation or to the corresponding like-dislike manipulation? Would either manipulation alone produce a result or do they work only in combination? If you obtain no results, is it perhaps because the two manipulations tend to cancel each other out? All these are questions you will need to consider when you tailor the present exercise for the use of your particular class. So, in formulating your experimental strategy, you will be weighing two considerations that, unfortunately, can work against each other: (1) whether to strengthen the manipulations to increase the probability that you will obtain differences between groups and (2) whether to simplify the manipulations so that you will be able to interpret unambiguously any differences obtained.

If you follow the script provided, you are likely to produce between-group differences in amount of attitude change. But you may not know exactly what caused the obtained differences. If you or your instructor decide to make the experiment more analytic by reducing the number of differences between the two experimental groups, simply cross out the appropriate portions of the script.

Data Analysis

The data from the Values Questionnaire will not be analyzed; these questionnaires were collected only to make the similarity manipulation seem plausible to the subjects. Items 1 and 2 on the Evaluation Questionnaire are essentially buffer or lead-in items; however, the groups may differ in their

answers to these questions. It may be interesting to speculate about why the one or two sentences that differentiate the various conditions should produce different answers to questions 1 and 2.

Questions 3 and 4 on the Evaluation Questionnaire are often called "take" measures or manipulation checks because they measure whether the manipulation "took" or worked. Is it the case that your manipulation led Positive subjects to see themselves as more similar to the author of the communication than do the Comparison or Negative subjects (Question 3)? Did Positive subjects also like the author better (Question 4)? What are the implications for your experiment if the groups do not differ on questions 3 and 4? It is possible that the operational definitions were inadequately conceived or executed (see the Introduction). Or perhaps questions 3 and 4 are insensitive or inappropriate measures for your particular experimental manipulation.

Three Data Sheets (pp. 159–163) are provided to facilitate computation of *t*-tests (Appendix A) between the three groups. The most appropriate statistic for comparing the three conditions is an analysis of variance; however, the simpler *t*-tests will yield an adequate analysis for your purposes.

Record responses to Q1 and Q2 on Data Sheet 1 (p. 159) and to Q3 and Q4 on Data Sheet 2 (p. 161). Q5, Q6, and Q7 are first added across to form a combined measure of attitude change and then summed on Data Sheet 3 (p. 163).

For each of the five dependent variables (Q1, Q2, Q3, Q4, and the combined attitude change measure), contrast the Positive condition mean with the Comparison mean using the *t*-test procedure (See Appendix A). A significant *t* indicates that the Positive manipulation facilitated attitude change relative to the Comparison group. Next, contrast the mean from the Negative condition with the Comparison mean to determine whether the negative manipulation inhibited attitude change relative to the Comparison group. Finally, contrast the Positive and Negative means. Discuss the pattern of results you have obtained. Does the analysis for questions 3 and 4 indicate that the experimental manipulations were effective? If not, are the attitude change results in the expected direction when only those subjects affected by the manipulation are considered?

References

Back, K. W. Influence through social communication. *Journal of Abnormal Social Psychology*, 1951, **46**, 9–23.

In this classic experiment, Back shows that three different techniques for inducing liking among group members all produce more social influence.

Berscheid, E., and Walster, E. *Interpersonal attraction*. Reading: Addison Wesley, 1969.

An excellent short paperback which deals with a number of topics related to interpersonal attraction.

Byrne, D. Attitudes and attraction. In L. Berkowitz (Ed.) *Advances in experimental social psychology, Vol. 4.* New York: Academic Press, 1969.

This is a summary of Byrne's research on similarity and liking. Byrne typically prepares bogus personality test results and presents them to subjects. The subjects indicate how much they are attracted to the people who supposedly took the personality tests. Subjects tend to like persons who are portrayed as similar to them. One of Byrne's studies is included as the Research Example following this Exercise.

Newcomb, T. M. Varieties of interpersonal attraction. In D. Cartwright and A. Zander (Eds.), *Group dynamics*, (2nd ed.). Evanston, Ill.: Row, Peterson & Co., 1960; Newcomb, T. M. *The acquaintance process*. New York: Holt, Rinehart, & Winston, 1961.

These references describe a field study of interpersonal attraction among members of a college living unit. Perceived and actual similarity were related to liking.

Sampson, E. E., and Insko, C. A. Cognitive consistency and performance in the autokinetic situation. *Journal of Abnormal and Social Psychology*, 1964, **68**, 184–192.

This study demonstrates that liked others have more social influence than disliked others.

Persuasive Communication

(The following is an example of a persuasive communication on raising the driver's age. Have someone else copy the communication over in their own handwriting on page 157. Don't do it yourself; the subject might recognize your handwriting sometime during the experiment.)

Actually, I feel fairly strongly on this issue. I don't think that 16- and 17-year-olds ought to be driving anywhere. First of all, they're the main reason that my insurance rates are so high. Even though people under 25 pay higher rates, insurance companies take a loss on teen-age drivers and it's guys like me who have to make up the difference.

For that matter, it's for their own good. An awful lot of teen-age drivers get killed—and they'd still be alive today if there were laws that made them wait until they were more mature before they could get behind the steering wheel. And it isn't just their lives that are at stake. They may take along some perfectly innocent and safe driver with them when they pass on a hill or do something else foolish.

The more I think of it the more sure I become. We should raise the legal driving age so that people have to be 18 before they can get a driver's license.

Experimenter's Summary Sheet and Verbatim Script

COMPARISON CONDITION

Hello, my name is ______________________________. My social psychology class is doing an experiment on the effectiveness of interpersonal communication. Would you be willing to read one of our specimen communications and tell us what you think about it? (If subject agrees:) Before you evaluate the communication we need to know a little bit about you, the "audience" for the communication. Would you fill out this brief questionnaire? [Hand subject *Attitude* Questionnaire (p. 156). Do *not* give values questionnaire or communication to subjects in this comparison condition.]

(After the subject has filled out the questionnaire, explain that he was in a control condition, and that you decided on a random basis to give him a questionnaire without first giving him a communication. Other subjects will receive a communication before they fill out the questionnaire. By contrasting the data from comparison subjects with that of subjects who received the communication, you will be able to determine whether the communication produced any attitude change.)

POSITIVE CONDITION

Hello, my name is ______________________________. My social psychology class is doing an experiment on the effectiveness of interpersonal communication. Would you be willing to read one of our specimen communications and tell us what you think about it? (If subject agrees:) Before you evaluate the communication, we need to know a little bit about you, the "audience" for the communication. Would you fill out this brief questionnaire? [Hand subject *Values* Questionnaire (p. 153).]

(Look questionnaire over when subject hands it back and pretend to compare it with another questionnaire or piece of paper.) This is very interesting. We asked the author of the communication you are about to read to fill out the same questionnaire, and in almost every case, he gave an answer similar to yours. The person who wrote the communication you are about to read and you are really very similar. I am sure that if you two ever got together, you'd like each other very much. I imagine you both come from similar backgrounds. (Pause.) Well here's the communication he wrote. I'd like you to read it over once and then I'll give you some questions to answer about it.

[Give the subject the handwritten communication. After he has read it, hand him the Evaluation Questionnaire (pp. 155-156). After the subject has

completed the questionnaire, explain to him that you copied the communication out of a textbook, and that you really don't know anything about the personality of the author. Explain to him that you are investigating the effect of similarity and/or attraction to attitude change. Take long enough to discuss the experiment with the subject so that he understands what you're trying to do and does not feel that he has been deceived without reason. You should then ask the subject not to discuss the experiment for a week (or for however long it will take the class to run the experiment). Explain to him that the whole class exercise will be ruined if word gets out about what you are trying to do.]

NEGATIVE CONDITION

Hello, my name is ________________________. My social psychology class is doing an experiment on the effectiveness of interpersonal communication. Would you be willing to read one of our specimen communications and tell us what you think about it? (If subject agrees:) Before you evaluate the communication, we need to know a little bit about you, the "audience" for the communication. Would you fill out this brief questionnaire? [Hand subject *Values* Questionnaire (p. 153).]

(Look questionnaire over and say:) This is very interesting. We asked the author of the communication you are about to read to fill out the same questionnaire, and in almost every case, he gave an answer different from yours. The person who wrote the communication you are about to read and you are really very different. I am sure that if you two ever got together, you'd dislike each other very much. I imagine you both come from different backgrounds. (Pause.) Well here's the communication he wrote. I'd like you to read it over once and then I'll give you some questions to answer about it.

[Give the subject the handwritten communication. After he has read it, hand him the Evaluation Questionnaire (pp. 155-156). After the subject has completed the questionnaire, explain the actual purpose of the exercise and pledge him to secrecy as described above in the Positive condition instructions.]

Values Questionnaire

(Circle a number for each statement.)

1. I think the federal government is too powerful.

agree strongly	agree slightly	don't know	disagree slightly	disagree strongly
1	2	3	4	5

2. I think that the new sexual behaviors are undermining the moral fiber of our country.

agree strongly	agree slightly	don't know	disagree slightly	disagree strongly
1	2	3	4	5

3. I think everybody should be entitled to a college education.

agree strongly	agree slightly	don't know	disagree slightly	disagree strongly
1	2	3	4	5

4. If more people recognized the value of hard work, this world would be a better place to live in.

agree strongly	agree slightly	don't know	disagree slightly	disagree strongly
1	2	3	4	5

5. There really isn't a job that a woman can't do as well as a man.

agree strongly	agree slightly	don't know	disagree slightly	disagree strongly
1	2	3	4	5

Evaluation Questionnaire

(Circle a number for each statement.)

1. I thought the ideas in the communication were:

expressed extremely clear	expressed moderately clear	expressed slightly clear	expressed neither clearly nor unclearly	expressed slightly unclear	expressed moderately unclear	expressed extremely unclear
1	2	3	4	5	6	7

2. I thought the author of the communication was probably:

extremely smart	moderately smart	slightly smart	neither smart nor dumb	slightly dumb	moderately dumb	extremely dumb
1	2	3	4	5	6	7

3. I thought the author of the communication was probably:

extremely like me	moderately like me	only slightly like me	neither like me nor different from me	slightly different from me	moderately different from me	extremely different from me
1	2	3	4	5	6	7

4. I think I would like the author of the communication:

very much	somewhat	only slightly	neither like nor dislike	dislike slightly	moderately dislike	dislike very much
1	2	3	4	5	6	7

As you might imagine, your own attitudes on this particular issue might affect how you rated the interpersonal communication on this topic. For this reason, we're asking you also to answer the questions on the reverse side, which give your own opinion to the topic of the communication.

Attitude Questionnaire

(Circle a number for each statement.)

5. No one under 18 should be allowed to drive:

agree definitely	agree	agree slightly	neither agree nor disagree	disagree slightly	disagree	disagree definitely
1	2	3	4	5	6	7

6. I would feel safer if 16- and 17-year-olds did not drive:

agree definitely	agree	agree slightly	neither agree nor disagree	disagree slightly	disagree	disagree definitely
1	2	3	4	5	6	7

7. People 16 and older are mature enough to drive:

agree definitely	agree	agree slightly	neither agree nor diagree	disagree slightly	disagree	disagree definitely
1	2	3	4	5	6	7

Persuasive Communication

This is a study on the effectiveness of interpersonal communication. Please use the space below to write a communication which gives your view on the question "Should the minimum age for obtaining a driver's license be raised to 18?" The message you write will be shown to other people who will be asked to evaluate it as an interpersonal communication.

Data Sheet 1
Evaluation Questionnaire

Subject	Comparison				Positive				Negative			
	Q3	$Q3^2$	Q4	$Q4^2$	Q3	$Q3^2$	Q4	$Q4^2$	Q3	$Q3^2$	Q4	$Q4^2$
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10.												
11.												
12.												
13.												
Sum X or X^2	Sum X	Sum X^2	Sum X	Sum X^2	Sum X	Sum X^2	Sum X	Sum X^2	Sum X	Sum X^2	Sum X	Sum X^2
	$\frac{\text{Sum } X}{N^a} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$	

[a] N equals the number of subjects within the treatment condition.

Data Sheet 2
Evaluation Questionnaire

Subject	Comparison				Positive				Negative			
	Q3	$Q3^2$	Q4	$Q4^2$	Q3	$Q3^2$	Q4	$Q4^2$	Q3	$Q3^2$	Q4	$Q4^2$
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10.												
11.												
12.												
13.												
Sum X or X^2	$\square$ Sum X	$\square$ Sum X^2	$\square$ Sum X	$\square$ Sum X^2	$\square$ Sum X	$\square$ Sum X^2	$\square$ Sum X	$\square$ Sum X^2	$\square$ Sum X	$\square$ Sum X^2	$\square$ Sum X	$\square$ Sum X^2
	$\frac{\text{Sum } X}{N^a} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$		$\frac{\text{Sum } X}{N} = \square$	

[a] N equals the number of subjects within the treatment condition.

Data Sheet 3
Attitude Questionnaire

Subject	Comparison					Positive					Negative				
	Q5	Q6	Q7	$(Q5+Q6+Q7)=x$	$(Q5+Q6+Q7)^2=x^2$	Q5	Q6	Q7	$(Q5+Q6+Q7)=x$	$(Q5+Q6+Q7)^2=x^2$	Q5	Q6	Q7	$(Q5+Q6+Q7)=x$	$(Q5+Q6+Q7)^2=x^2$
1.															
2.															
3.															
4.															
5.															
6.															
7.															
8.															
9.															
10.															
11.															
12.															
13.															
Sum X or X^2	Comparison $\boxed{}$ $\boxed{}$ Sum X Sum X^2 $\frac{\text{Sum } X}{N^a} =$ $\boxed{}$ Comparison mean					Positive $\boxed{}$ $\boxed{}$ Sum X Sum X^2 $\frac{\text{Sum } X}{N} =$ $\boxed{}$ Positive mean					Negative $\boxed{}$ $\boxed{}$ Sum X Sum X $\frac{\text{Sum } X}{N} =$ $\boxed{}$ Negative mean				

[a] N equals the number of subjects within the treatment condition.

Research Example

Interpersonal Attraction and Attitude Similarity[1]

DONN BYRNE

In investigating the direction and the strength of the affect engendered between the two participants in a dyad, we may arrange the expressed feelings of each individual along a continuum ranging from strongly positive to strongly negative. The accurate prediction of interpersonal attraction and repulsion in such relationships will undoubtedly require that we secure knowledge about several classes of independent variables.

Probably the most obvious and also best documented variable is that of propinquity. Studies in a wide variety of settings have shown that physical and functional distance influence interaction and interpersonal attraction (Byrne, in press). Once the environmental situation permits or encourages interaction, affiliation need should be helpful in predicting individual differences in interpersonal behavior (Atkinson, Heyns, & Veroff, 1954; Schachter, 1959). A third class of variables consists of the overt stimulus properties of each individual to which other individuals would be expected to respond on the basis of generalization from previous interpersonal interactions.

Once interaction has begun, reciprocal reward and punishment is proposed as the crucial determining factor. It has been suggested (Newcomb, 1956) that attraction between persons is a function of the extent to which reciprocal rewards are present in their interaction; perhaps dislike is a function of reciprocal punishments. A special subclass of this variable would be perceived similarity and dissimilarity of the attitudes of two individuals. It can be assumed that persons in our culture have well established

SOURCE. *Journal of Abnormal and Social Psychology*, 1961, **62**(3), 713–715.

[1] A portion of this paper was read at the meetings of the Southwestern Psychological Associaton, Galveston, 1960.

This investigation was supported in part by a research grant (EF-140) from the University of Texas Research Institute.

learned drives to be logical and to make a correct report of the environment. Those who seem deficient in this respect are generally categorized as being uninformed, of low intelligence, immoral, and/or as being out of contact with reality. It is primarily through consensual validation that we determine whether we or anyone else is logical or correct in interpreting environmental events. Hence, any time that another person offers us validation by indicating that his percepts and concepts are congruent with ours, it constitutes a rewarding interaction and, hence, one element in forming a positive relationship. Any time that another person indicates dissimilarity between our two notions, it constitutes a punishing interaction and thus one element in forming a negative relationship. Disagreement raises the unpleasant possibility that we are to some degree stupid, uninformed, immoral, or insane. An alternative possibility is that it is the other person who is deficient in one or more of these characteristics. Probably other variables, such as the importance of the issue to each individual, contribute to the effect.

A number of studies have found greater similarity among friends than among nonfriends with respect to a variety of issues (Bonney, 1946; Loomis, 1946; Newcomb, 1956; Precker, 1952; Richardson, 1940; Winslow, 1937). A few studies of a more experimental nature also support the notion of a relationship between attitude similarity and interpersonal attraction (Jones & Daugherty, 1959; Smith, 1957).

In order to test the proposition that the effect of attitude similarity is a causative one and to test some implications arising from the preceding speculations about the reason for the effect, it was hypothesized that (*a*) a stranger who is known to have attitudes similar to those of the subject is better liked than a stranger with attitudes dissimilar to those of the subject, (*b*) a stranger who is known to have attitudes similar to those of the subject is judged to be more intelligent, better informed, more moral, and better adjusted than a stranger with attitudes dissimilar to those of the subject, and (*c*) a stranger who is known to have similar attitudes on issues important to the subject and dissimilar attitudes on unimportant issues is better liked and is evaluated more positively on the other four variables than a stranger for whom the reverse is true.

Method

ATTITUDE MEASURE[2]

On the basis of a pilot study, 26 issues were selected for inclusion in an attitude and opinion scale. Each issue was presented in a seven-point scale. The issues ranged from those thought to be extremely important by the pilot subjects (e.g., integration, God, premarital sex

[2] The attitude measure, response frequencies, and the rating scales for interpersonal attraction and evaluation have been deposited with the American Documentation Institute. Order Document No. 6770 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance $1.75 for microfilm or $2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

relations) to those considered to be of minor importance (e.g., western movies and television programs, classical music, politics).

PROCEDURE

The attitude scale was administered to 64 students (36 male, 28 female) enrolled in an introductory psychology course at the University of Texas. Response heterogeneity differed from item to item, but there was moderately wide diversity of opinion among the 64 subjects. After filling out the attitude scale, the subjects were asked to indicate which they believed to be the 13 most important and 13 least important issues.

Two weeks later they were falsely informed that the attitude scale had been given as part of a study in interpersonal prediction. They were told that individuals in another class had been given the same scale that they took, students in the two classes were matched on the basis of sex, and they were to be given each other's tests (name removed) in order to determine how much they could learn about one another from this information alone.

Actually the questionnaire they received at this time was a fake one made up by the experimenter. The subjects had been randomly divided into four groups; one group received attitude scales filled out exactly the same as theirs had been, one received scales with exactly opposite views expressed, one received scales with similar opinions on the most important issues and dissimilar on the least important, and the fourth received scales with similar opinions on the least important issues and dissimilar opinions on the most important. The four groups of subjects did not differ significantly in their initial responses to any of the 26 issues.

INTERPERSONAL ATTRACTION AND EVALUATIONS

A rating scale was used as the measure of interpersonal attraction and evaluation with each dependent variable represented in a seven-point scale. As a measure of interpersonal attraction, subjects were asked to indicate how well they felt they would like this person and whether they believed they would enjoy working with him (or her) as a partner in an experiment. Four scales dealt with evaluation; the subjects were asked for their judgments as to the other sutdent's intelligence, knowledge of current events, morality, and adjustment.

Results

FIRST HYPOTHESIS

Table 1 shows the comparisons of the two groups on each of the dependent variables. The first hypothesis was overwhelmingly confirmed for each of the two attraction scales. The group with attitude scales filled out the same as their own (SA) indicated significantly more positive feelings toward the "stranger" than did the group which received scales indicating dissimilar attitudes (DA). Each difference was significant at less than the .001 level.[3]

[3] Notations concerning significance levels, such as less than the .001 level are explained in the introduction, pp. 15-16.

Table 1

Comparison of the Similar Attitude (SA) and Dissimilar Attitude (DA) Groups on Interpersonal Attraction and Evaluation

	SA (*N* = 17)		*DA* (*N* = 17)		*D*	*t*	*df*	*p*
	M	*SD*	*M*	*SD*				
Personal Feelings	6.53	.50	1.76	.73	4.77	21.46	32	$<.001$
Desirability as Work Partner	6.47	.50	2.65	1.88	3.82	7.88	32	$<.001$
Intelligence	5.65	.68	3.06	.87	2.59	9.37	32	$<.001$
Knowledge of Current Events	4.65	1.14	2.65	.91	2.00	5.51	32	$<.001$
Morality	5.76	.73	3.47	2.09	2.29	4.14	32	$<.001$
Adjustment	6.00	.84	2.71	1.13	3.29	9.36	32	$<.001$

Table 2

Comparison of the Similar on Important Attitudes (SIA) and Similar on Unimportant Attitudes (SUA) Groups on Interpersonal Attraction and Evaluation

	SIA (*N* = 15)		*SUA* (*N* = 15)		*D*	*t*	*df*	*p*
	M	*SD*	*M*	*SD*				
Personal Feelings	4.20	1.51	2.60	1.20	1.60	3.10	28	$<.01$
Desirability as Work Partner	4.27	1.44	3.33	1.40	.94	1.76	28	*ns*
Intelligence	4.13	.62	3.73	1.34	.40	1.01	28	*ns*
Knowledge of Current Events	3.60	.95	3.53	.96	.07	.19	28	*ns*
Morality	5.33	1.25	3.33	1.66	2.00	3.60	28	$<.01$
Adjustment	4.07	1.57	2.93	1.18	1.14	2.17	28	$<.05$

SECOND HYPOTHESIS

As is indicated in Table 1, the second hypothesis was also confirmed. The SA group rated the "stranger" significantly higher than did the DA group on intelligence, knowledge of current events, morality, and adjustment. Again, each difference reached a level of significance beyond the .001 level.

THIRD HYPOTHESIS

The third hypothesis, concerning the influence of important vs. unimportant issues, was only partially confirmed. As is shown in Table 2, the Similar on Important Attitudes Group (SIA) rated the "stranger" significantly more positively than did the Similar on Unimportant Attitudes Group (SUA) with respect to their personal feelings about him, his morality, and his adjustment. On the other three variables, the two groups did not differ.

Discussion

The experimental confirmation of the first two hypotheses is very encouraging for further research designed to investigate other aspects of the relationship

between interpersonal attraction and attitude similarity. It should be possible now to study the effect of attitude differences less extreme than those in the present study and to combine this variable with the others that influence interpersonal attraction in order to determine interaction effects.

Because of the fact that this group of subjects showed a degree of homogeneity of opinion on some of the attitude items, a possible alternative interpretation is that they were responding negatively to unusual and deviant beliefs rather than to disagreement per se. On 19 of the the 26 issues it was possible for a subject to fall in the deviant one fourth of the group by expressing a positive or a negative opinion. The range among the subjects was from no deviant attitudes to nine; they were divided into high and low subgroups on the basis of this score. Since these "conforming" and "deviant" subgroups did not differ from one another in responding to strangers with similar vs. different attitudes, there is no evidence to support this other interpretation of the results.

The partial failure of the third hypothesis led to a comparison of all four groups on each of the dependent variables. The results suggest that the Personal Feelings scale is the most sensitive measure of interpersonal attraction. With the other five interpersonal judgment scales, additional factors apparently contribute to the variance.

References

Atkinson, J. W., Heyns, R. W., & Veroff, J. The effect of experimental arousal of the affiliation motive on thematic apperception. *J. abnorm. soc. Psychol.,* 1954, **49**, 405–410.

Bonney, M. E. A sociometric study of the relationship of some factors to mutual friendships on the elementary, secondary, and college levels. *Sociometry*, 1946, **9**, 21–47.

Byrne, D. The influence of propinquity and oportunities for interaction on classroom relationships. *Hum. Relat.*, in press.

Jones, E. E., & Daugherty, B. N. Political orientation and the perceptual effects of an anticipated interaction. *J. abnorm. soc. Psychol.*, 1959, **59**, 340–349.

Loomis, C. P. Political and occupational cleavages in a Hanoverian village, Germany: A sociometric study. *Sociometry*, 1946, **9**, 316–333.

Newcomb, T. M. The prediction of interpersonal attraction. *Amer. Psychologist,* 1956, **11**, 575–586.

Precker, J. A. Similarity of valuings as a factor in selection. *J. abnorm. soc. Psychol.*, 1952, **47**, 406–414.

Richardson, Helen M. Community of values as a factor in friendships of college and adult women. *J. soc. Psychol.*, 1940, **11**, 303–312.

Schachter, S. *The psychology of affiliation.* Stanford: Stanford Univer. Press, 1959.

Smith, A. J. Similarity of values and its relation to acceptance and the projection of similarity. *J. Psychol.*, 1957, **43**, 251–260.

Winslow, C. N. A study of the extent of agreement between friends' opinions and their ability to estimate the opinions of each other. *J. soc. Psychol.,* 1937, **8**, 433–442.

Exercise 7

Spatial Distance

Most of us are aware that spatial arrangements of buildings, furniture, and people can have considerable impact on our moods and feelings, but only recently have behavioral scientists begun systematic investigation of the effects of these environmental variables.

In the past few years, psychologists, anthropologists, and architects have studied both the antecedents and consequences of perhaps the most interesting spatial variable—distance among people. For example, Hall (1966) has found that the "appropriate" distance for conversation varies from culture to culture, for example, two Arabs stand much closer to each other when talking than do two Americans.

Even within the same culture, different kinds of people may prefer different distances. Highly anxious individuals tend to stand farther away from others than less anxious subjects (King, 1966). Likewise, introverts distance themselves further from others than do extroverts (Sommer, 1969). Conversely, some studies have been concerned with the effects of spacing on people's attitudes toward others. In one such study, Patterson and Sechrest (1970) found that individuals who sat over two feet away from another person were judged less favorably by that person than those who sat at a closer distance.

In this exercise, you will investigate the use of physical spacing as a diagnostic tool. The assumption is that the comfortableness experienced by a person with another may be reflected by their physical proximity. This exercise then is designed to demonstrate that even a mild state of distress or anxiety will cause the threatened person to increase distance between himself and the threatening person.

METHOD

Subjects and Collaborators

Two groups consisting of 10 to 15 subjects each are necessary. An experimental group will be told that they will be interviewed by someone who is likely to be hostile, abrupt, and disagreeable (anxiety condition). Control or comparison subjects will not be given any description of the interviewer. You may wish to run a third group, in which the subjects are informed that the interviewer is mild-mannered, gentle, and well-liked (comfortable condition). The inclusion of this third group will allow the experimenter to determine not only whether social uneasiness will increase the spatial distance between a social pair, but also whether a comfortable social interaction will move people closer together. Since there may be differences between the sexes in distancing behavior, it may be desirable to match the groups on gender. Therefore, groups can be composed of only female or only male subjects, or equal numbers of both sexes can be assigned to each group.

One collaborator will be needed to serve as the "interviewer." Use the same collaborator throughout the experiment. Since the interviewer might wittingly or unwittingly affect the behavior of the subject if he is aware of the treatment condition, he should not be told how he has been described to any particular subject; he should be "blind" to the experimental treatment given the subject.

Research Site

Virtually any research site that provides adequate privacy will be suitable. Once a room or office is selected, all subjects should be tested there. Since the collaborator will be measuring spatial distances between himself and the subject, it would be very convenient to have a "built-in" measuring device for this purpose. For example, you might use a room with 9 × 9-inch tile flooring. In such a setting the interviewer can use the number of tiles separating his chair (or feet) from the interviewee as a measure of distance. Wall patterns with vertical stripes or wood paneling may also be used for the purpose of measurement.

Procedure

After locating an appropriate experimental setting, station the interviewer at some fixed point in the room. In the Anxiety condition, instructions are designed to make subjects anxious. Subjects in the Comparison condition will receive nonanxious or "neutral" instructions. The exact script and

specific procedures are presented in the Experimenter's Summary Sheet and Script. As mentioned above, you may want to include a Comfortable condition in which subjects receive instructions designed to put them at ease with the interviewer.

Following the instructions, direct the subject to the door and see that he enters. Do not introduce him to the interviewer, and be sure that the interviewer does not know which instructions the subject has received.

The interview room should include two chairs, one for the interviewer, the other for the subject. The subject's chair should be faced away from the interviewer, and the distance between the two chairs should exceed seven feet since this is the maximum distance usually found comfortable for seated conversation. When the subject enters the room, the interviewer should instruct him to "take a chair." At this point, the subject will most likely turn his chair toward the interviewer (if he doesn't do this, the interviewer should ask him to do so). As he turns the chair, the subject has the option of drawing it closer to the interviewer. The distance between the subject's chair and the interviewer prior to the first question of the interview is the measure of spatial distance. The measure is recorded at the beginning of the interview because the subject may move his chair as he becomes more or less comfortable. An example of an appropriate interview topic is given on page 183. Because the interview may affect the mood of the subject, it is important that the interviewer try to treat each subject in as similar a manner as possible. The interviewer should keep in mind that he should not be more friendly to one subject than another, that the questions he poses should be worded identically when presented to different subjects, and that he should treat the subject with courtesy at all times. You may find it necessary to instruct your interviewer concerning appropriate behavior within a structured interview. At the end of the interview, the interviewer should thank each subject and dismiss him. If the subject has any questions, the interviewer should direct him to you, the experimenter. Following the interview, meet each subject outside of the room, and administer the Post-Experimental Questionnaire (page 187) as a check on the experimental manipulations.

DATA ANALYSIS

Distance

Record the initial separation distances on Distance Data Sheet 1 (page 177). The Data Sheet is provided to record subjects' behavior. Scores for Anxious and Comparison subjects should be recorded in columns 1 to 6. If

you ran a Comfortable group, their data should be entered in columns 7, 8, and 9. Compare each pair of means with a *t*-test (Appendix A).[1] Did the Anxious subjects station themselves farther away from the interviewer than the Comparison subjects?

Manipulation Check and Take Measures

If the groups did not differ, could this be attributable to a weak experimental manipulation? Check your Post-Experimental Questionnaire data to assess this possibility. Item 2 on the post-experimental questionnaire is directly relevant to this hypothesis. Did the subjects in the experimental group indicate greater initial nervousness during the interview than control subjects? To formally assess this possibility, score each response on a five-point scale with one end of the scale indicating high nervousness (Choice e = five points), the other end, no nervousness (Choice a = one point). Enter the data from the Post-Experimental Questionnaire on Data Sheet 2, page 179. Compare the means of experimental and comparison groups by *t*-tests. Did your experimental manipulations, in fact, produce differences between the groups in the average level reported anxiety?

Correlation Between Anxiety and Spatial Distance

We began with the hypothesis that the moods or mental states of individuals would influence their spatial distancing behavior; was this true for your subjects? Whether your *manipulation* of discomfort increased the interaction distance, you may combine the data from all subjects and assess the relationship between *reported* discomfort and distancing acts. Score each subject's response on the three Post-Experimental Questionnaire items as either 1 (most at ease), 2, 3, 4, or 5 (most discomfort). Add these scores together for each respondent. Thus, the anxiety score for each subject may range from 3 to 15. On Data Sheet 3 (page 181) *rank* subjects on (1) their actual distance measure as indicated on Data Sheet 1 and (2) their reported anxiety score on the questionnaire items as indicated on Data Sheet 2. Now you can determine the correlation between reported anxiety and initial distancing response. Using the rank-order method (Data Sheet 3 and Appendix C), correlate the subjects' self-reports of anxiety with their distance scores. Was this correlation significant? (See the Introduction, page 15 for a discussion of significance.)

[1] If you run three groups, there will be three possible comparisons among the three group means. Strictly speaking, the data from the three groups should be analyzed with a statistical technique called *analysis of variance* rather than multiple comparisons with *t*-test. With as few as three groups, the simple *t*-test procedure is a close approximation to the more correct analysis of variance.

References

Hall, E. T. *The hidden dimension*. New York: Doubleday, 1966.

An anthropologist's account of the effects of spatial arrangements on human behavior in a variety of cultural settings.

King, M. G. Interpersonal relations in preschool children and average approach distance. *Journal of Genetic Psychology,* 1966, **108**, 109–116. See text.

Patterson, M. Spatial factors in social interaction. *Human Relations*, 1968, **21**, 351–361.

This article reviews studies pertaining directly to this exercise. Studies relevant to social interaction are extensively covered. See text.

Patterson, M. L., and Sechrest, L. B. Impression formation and interpersonal distance. *Journal of Personality,* 1970, **38**, 161–166. See text.

Sommer, R. Small group ecology. *Psychology Bulletin,* 1967, **67**, 145–152.

This article reviews studies concerning the "arrangement of individuals in small groups. . . ." Experimental studies of spatial arrangements and leadership, task and personality variables are reviewed. One of Sommer's studies is appended as a Research Example.

Data Sheet 1
Distance Scores

Column Numbers											
1		2	3	4		5	6	7		8	9
Subject number and sex Anxious Ss		Initial seating distance Anxious Ss	(Initial seating distance Anxious Ss)^2	Subject number and sex Comparison Ss		Initial seating distance Comparison Ss	(Initial seating distance Comparison Ss)^2	Subject number and sex Comfortable Ss		Initial seating distance Comfortable Ss	(Initial seating distance Comfortable Ss)^2
1				16				31			
2				17				32			
3				18				33			
4				19				34			
5				20				35			
6				21				36			
7				22				37			
8				23				38			
9				24				39			
10				25				40			
11				26				41			
12				27				42			
13				28				43			
14				29				44			
15				30				45			

Col 2	Col 3	Col 5	Col 6	Col 8	Col 9
Sum Anxious Ss	Sum X^2	Sum Comparison Ss	Sum X^2	Sum Comfortable Ss	Sum X^2

$$\text{Mean} = \frac{\text{Sum}}{N}$$

Mean Anxious Ss	Mean Comparison	Mean Comfortable Ss

Data Sheet 2

Reported Anxiety Scores (from Post-Experimental Questionnaires)

Column Numbers											
1		2	3	4		5	6	7		8	9
Subject number and sex Anxious *Ss*		Initial seating distance Anxious *Ss*	(Initial seating distance Anxious *Ss*)^2	Subject number and sex Comparison *Ss*		Initial seating distance Comparison *Ss*	(Initial seating distance Comparison *Ss*)^2	Subject number and sex Comfortable *Ss*		Initial seating distance Comfortable *Ss*	(Initial seating distance Comfortable *Ss*)^2
1				16				31			
2				17				32			
3				18				33			
4				19				34			
5				20				35			
6				21				36			
7				22				37			
8				23				38			
9				24				39			
10				25				40			
11				26				41			
12				27				42			
13				28				43			
14				29				44			
15				30				45			

	Col 2	Col 3	Col 5	Col 6	Col 8	Col 9
	Sum Anxious Ss	Sum X^2	Sum Comparison *Ss*	Sum X^2	Sum Comfortable *Ss*	Sum X^2
Mean = $\frac{\text{Sum}}{N}$	Mean Anxious *Ss*		Mean Comparison		Mean Comfortable *Ss*	

Data Sheet 3
Correlation between Self Report of Anxiety and Distance

Subject numbers for all subjects	Rank on Distance Scores	Rank on Anxiety	Difference Between Columns (2) and (3)
1			
2			
3			
4			
5			
6			
7			
8			
9			
10			
11			
12			
13			
14			
15			
16			
17			
18			
19			
20			
21			
22			
23			
24			
25			
26			
27			
28			
29			
30			

Data Sheet 3 continued on p. 182

Subject numbers for all subjects	Rank on Distance Scores	Rank on Anxiety	Difference Between Columns (2) and (3)
31			
32			
33			
34			
35			
36			
37			
38			
39			
40			
41			
42			
43			
44			
45			

Refer to Appendix C to compute correlation.

Questionnaire and Data Sheet on Legalizing Marijuana

Initial distance from interviewer ____________ Condition: (Circle one)

Final distance from interviewer ____________

Subject's name ____________________

Date ____________

Interviewer ____________________

Anxiety

Neutral

Ease

1. Do you think that smoking marijuana is harmful to your health?

   Yes   No   (why?)

2. Do you think that smoking marijuana leads the smoker to using drugs like heroin?

   Yes   No   (why?)

3. Should use of marijuana be legal?

   Yes   No   (why?)

4. If marijuana was legalized, should there be a legal age limit for its use?

   Yes   No   (If yes, what age should be specified?)

5. Have you ever been in the company of someone smoking marijuana?

   Yes   No   (If yes, did this affect your views on this subject?)

Experimenter's Summary Sheet and Script for Spatial Distance Exercise

ANXIETY INDUCTION SCRIPT

"I am doing a survey for a social psychology class and am interested in obtaining your views concerning the legalization of marijuana. I have a few questions that I would like you to answer about this subject. It will take only a few minutes of your time. Unforunately, I have to see some other respondents so another person will be interviewing you. He is in this room. Between you and me, I am not too happy with this interviewer. The others whom he has interviewed have reported having a rather unpleasant experience. They say he is abrupt, rude, and rather aggressive; in general, he makes them feel quite uncomfortable. He may well make you feel the same way. He's in this room (nod toward interview room); so go ahead in. I'll try to see you after the interview."

NEUTRAL OR CONTROL SCRIPT

"I am doing a survey for a social psychology class and am interested in obtaining your views concerning the legalization of marijuana. I have a few questions that I would like you to answer about this subject. It will take only a few minutes of your time. Unfortunately, I have to see some other respondents so another person will be interviewing you. He's in this room (nod toward interview room); so go ahead in. I'll try to see you after the interview."

COMFORTABLE INDUCTION SCRIPT

If you decide to run a third group of subjects to assess the impact of comfortableness upon distancing, the following instructions can be used:

"I am doing a survey for a social psychology class and am interested in obtaining your views concerning the legalization of marijuana. I have a few questions that I would like you to answer about this subject. It will take only a few minutes of your time. Unfortunately, I have to see some other respondents so another person will be interviewing you. He is in this room. Between you and me, I am rather happy with this interviewer. The others whom he has interviewed have reported having a rather pleasant experience. They say that he is polite and rather nice and in general makes them feel quite comfortable. He may well make you feel the same way. He's in this room (nod toward interview room); so go ahead in. I'll try to see you after the interview."

Post-Experimental Questionnaire

Since many people feel either quite comfortable or uncomfortable in a new situation, we are interested in how the interviewer's manner and personality may have affected your mood. Please check the alternative that best describes your feelings about the interview. Your comments will remain confidential.

1. In general, did the interview make you feel

   (a) extremely comfortable.
   (b) somewhat comfortable.
   (c) both comfortable and uneasy.
   (d) somewhat uneasy.
   (e) extremely uneasy.

2. Upon entering the interview situation, how would you describe your mood?

   (a) I was quite comfortable—no trace of uneasiness.
   (b) I was somewhat comfortable.
   (c) I was only slightly uneasy.
   (d) I was somewhat uneasy but not excessively nervous.
   (e) Actually, I was quite nervous about the whole procedure and/or the interviewer.

3. Did your feelings about the interviewer change as you interacted with him?

   (a) He made me feel considerably more comfortable.
   (b) He made me feel somewhat more comfortable.
   (c) My feelings didn't change at all.
   (d) He made me feel somewhat more uneasy.
   (e) He made me feel considerably more uneasy.

Research Example

Territorial Defense and the Good Neighbor[1]

ROBERT SOMMER AND
FRANKLIN D. BECKER

A series of questionnaire and experimental studies was designed to explore how people mark out and defend space in public areas. The use of space is affected by instructions to defend actively the area or retreat, by room density, and by the location of walls, doors, and other physical barriers. Under light population pressure, most markers are capable of reserving space in a public area, but more personal markers have the greatest effect. As room density increases, the effect of the marker is seen in delaying occupancy of the area and in holding onto a smaller subarea within the larger space. Neighbors play an important part in legitimizing a system of space ownership.

The concept of human territoriality is receiving increased attention. In addition to the popular books by Ardrey (1961, 1966), a number of social scientists have become impressed with the utility of the concept (Altman and Haythorn, 1967; Esser et al., 1965; Hall, 1966; Lipman, 1967; Lyman & Scott, 1967). Hediger (1950) defined a territory as "an area which is first rendered distinctive by its owner in a particular way and, secondly, is defended by the owner." When the term is used by social scientists to refer to human behavior, there is no implication that the underlying mechanisms are identical to those described in animal research. The major components of Hediger's definition are *personalization* and *defense*. Roos (1968) uses the term *range* as the total area an individual traverses, *territory* as the area he defends, *core area* as the area he preponderantly occupies, and *home* as the area in which he sleeps. Goffman (1963) makes the further distinction between a territory and a *jurisdiction,* such as that exercised by a janitor sweeping the floor of an office and keeping other people away. Territories are defended on two grounds, "you keep off" and "this space is mine." Jurisdictions are controlled only on the

SOURCE. *Journal of Personality and Social Psychology*, 1969, **11**, (2), 85–92.

[1]The authors are grateful to Harriet Becker, Martha Connell, Ann Gibbs, Lee Mohr, Tighe O'Hanrahan, Pamela Pearce, Ralph Requa, Sally Robison, and Nancy Russo for their assistance. Requests for reprints should be sent to Robert Sommer, Department of Psychology, University of California, Davis, California 95616.

former ground; no claim of ownership, no matter how transitory, is made.

In a previous study, the reactions to staged spatial invasions were investigated (Felipe & Sommer, 1966). There was no single reaction to a person coming too close; some people averted their heads and placed an elbow between themselves and the intruder, others treated him as a nonperson, while still others left the area when he came too close. The range of defensive gestures, postures, and acts suggested that a systematic study of defensive procedures would contribute materially to our knowledge of human spatial behavior. Following the tradition of ecological research, the studies would be undertaken in naturally occurring environments.

Questionnaire Studies

During previous observations of library study halls Sommer (1967) was impressed by the heavy concentration of readers at the side-end chairs. Interviewing made it clear that students believed that it was polite to sit at an end chair. Someone who sat, for example, at a center chair of an empty six-chair table (three chairs on each side) was considered to be "hogging the table." There appeared to be two styles by which students gained privacy in the library areas. One method was avoidance, to sit as far away from other people as one could. The other method was offensive ownership of the entire area. To study the two methods of gaining privacy, a brief questionnaire was constructed which presented the student with table diagrams containing 6, 8, and 10 chairs, respectively (Sommer, 1967). Two forms to the questionnaire were distributed randomly within a class of 45 students. Twenty-four students received avoidance instructions: "If you wanted to be as far as possible from the distraction of other people, where would you sit at the table?" Twenty-one other students in the same class were shown the same diagrams and given the offensive display instructions: "If you wanted to have the table to yourself, where would you sit to discourage anyone else from occupying it?" Even though both sets of instructions were aimed at insuring privacy, the two tactics produced a striking difference in seats chosen. Those students who wanted to sit by themselves as far as possible from other people overwhelmingly chose the *end* chairs at the table, while those students who wanted to keep other people away from the table almost unanimously chose the *middle* chair.

When the findings were discussed with architect James Marston Fitch, his first question concerned the location of the door in regards to the table. This seemed a good question, since the preferred location for retreat or active defense should be guided by the path the invaders would take or by the most accessible escape route. The previous diagrams had depicted only a table and chairs, so it seemed necessary to undertake another study in which the entrance to the room was indicated. This conception of the study suggested that additional information could be obtained on the ecology of retreat and active defense by varying the location of walls and aisles and the table size.

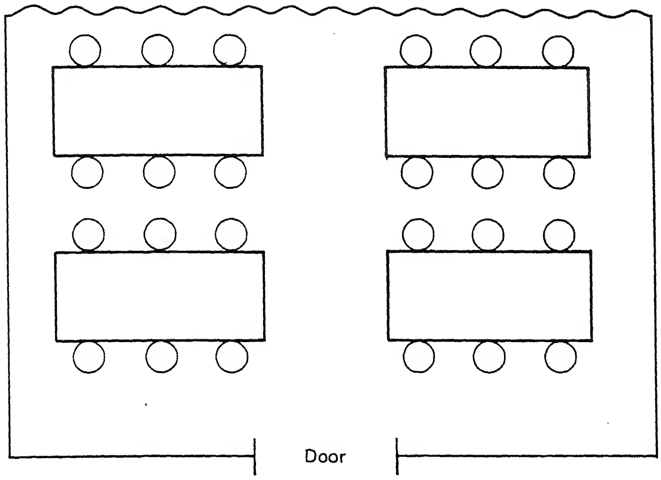


Figure 1. Arrangement of tables and chairs in Form G.

METHOD

The present study involved four diagrams, each one drawn on a separate 8½ X 11-inch sheet.

Form G showed eight rectangular six-chair tables, with a large aisle down the center and two smaller aisles along the walls. (See Figure 1.)

Form H was the same as Form G, only the tables were set against the wall and the center aisle was wider.

Form J was a hybrid of G and H, with the right row of tables against the wall and the left row of tables away from the wall.

Form I contained one row of four-chair tables and one row of eight-chair tables, with aisles in the center and along both walls.

Four different sets of instructions were used with the forms (two defense styles and two densities), but any single subject received only one set. One form asked the subject where he would sit if he wanted to be by himself and away from other people—the retreat instructions. The other form asked where he would sit if he wanted to keep other people away from the table—the active defense instructions. In each case, the prospective room density was also indicated. On half the questionnaires, it was stated that room density was likely to be low throughout the day and very few people would be using the room, while remaining subjects were told that room density was likely to be high and many people would be using the room. All the instructions described the room as a study hall such as that already existing in the campus library, and the respondent was informed that he was the first occupant in the room, so he could take any seat he wanted. Booklets containing some combination of instructional set (Defense Style X Room Density) and two diagrams in random order were passed out randomly among 280 students in introductory psychology classes.

RESULTS

Hypothesis 1 stated that during the retreat condition people gravitate to the end chair closest to the wall. During the active defense condition they make greater use of the center and aisle chairs. Hypothesis 1 was confirmed beyond the .01 level.[2] During the retreat conditions 76% of the subjects occupied a wall chair compared to 48% during the active defense condition.

Hypothesis 2 stated that with the retreat instructions the subjects face away from the door, while they face towards the door with the active defense instructions. The data disclose a preference in all conditions for a subject to sit with his back to the door—60% of the subjects faced away from the door compared to 40% who faced towards it. However, the results were still in the predicted direction since 44% of the subjects in the active defense condition faced the door compared to 36% in the retreat condition ($p < .05$).

Although the authors had imagined that the use of different-sized tables and the variation in wall placement would influence seating patterns, specific hypotheses had not been formulated. In all conditions there was a marked preference for chairs towards the rear of the room. Overall, 79% selected chairs in the rear half of the room. However, occupancy of the rear was significantly higher with the retreat instructions under high room density than in any of the other conditions ($p < .05$). There was also a highly significant preference for the four-chair tables when they were paired with the eight-chair tables, with 73% selecting a small table compared with 27% selecting a large table. There was a slight trend in the active defense condition to make greater use of the small tables, but this was not statistically significant.

When tables against the wall were paired with tables with aisles on both sides, 62% of the subjects selected a table against the wall compared to 38% who chose a table with aisles on both sides ($p < .001$). As an independent variable, description of the projected room density as high or low made very little difference in where people sat. However, density interacted with the defense instructions on several of the tabulations. With high density *and* retreat instructions, there was significantly greater use of (*a*) the rear half of the room, (*b*) a wall compared to an aisle table, and (*c*) the chair closest to the wall. In essence, the attribution of high room density increased the degree of physical retreat. It had no observable effects on the active defense conditions.

The results make it clear that room dimension and the location of barriers must be considered if we are to understand the ecology of spatial defense. In a library reading room, the best chair for retreat is at the rear, facing away from the door, next to a wall, and at a small table if one is available. Distance from the door protects the person against people simply walking by as well as lazy intruders who are more likely to sit in the first available chair; facing away from the door tends to minimize distraction and also displays an antipathy toward social intercourse; a wall table protects a person's entire left (or right) side; an a small table reduces the num-

[2] Notations such as .01 level are explained in the Introduction, pp. 15-16.

ber of invaders in close proximity. At this point the authors felt they had derived many useful hypotheses from the questionnaire data which they wanted to extend using an experimental approach under natural conditions. The first experimental studies took place in two soda fountains, and the remainder took place in library areas.

Experimental Studies

Most territories are marked and bounded in some clear way. In the animal kingdom, markers may be auditory (bird song), olfactory (glandular secretions by deer), or visual (bear-claw marks on a tree). Since humans rely almost exclusively on visual markers, the authors decided to test the strength of various markers ranging from the physical presence of a person to impersonal artifacts.

STUDY 1

The first study took place in a popular soda fountain on campus. The soda fountain was located in a converted office building which still contained a number of small rooms. Patrons would obtain their refreshments at a central counter and then repair to one of the smaller rooms to eat and chat informally. Prior to the study, the authors had been struck by the sight of students walking up and down the corridor looking for an empty room. One of the small rooms which contained three square tables, each surrounded by four chairs, was used for the study. A 20-year-old girl who appeared to be studying stationed herself at a table facing the door. On other occasions during the same hours she stationed herself down the hall so she could observe who entered the experimental room. A session took place only when the room was unoccupied at the outset.

If an all-or-none criterion of room occupancy is applied, the experimenter's defense was not very successful. During only 1 of the 10 experimental sessions was she able to keep the entire room to herself. The average length of time before the room was occupied during the experimental sessions was 5.8 minutes compared to 2.6 minutes during the control sessions, but the difference was not statistically reliable. Although the experimenter was unable to keep the room to herself, she was able to protect the table at which she studied. The remaining three seats were occupied only once during the experimental sessions compared to 13 occupancies during the control sessions ($p < .01$). It seems clear that territorial defense in a public area is not an all-or-none affair. The defender's presence may be seen in a delay in occupancy rather than an absence of invaders and in the avoidance of a subarea within the larger area.

STUDY 2

The next study took place in a more traditional open-plan soda fountain and, instead of the physical presence of the experimenter, three sorts of objects were used as territorial markers—a sandwich wrapped in cellophane, a sweater draped over a chair, and two paperback books stacked on the table. In each case the experimenter located two adjacent empty tables and arbitrarily placed a

marker on one with the other as a control. Seating himself some distance away, he was able to record the duration of time before each table was occupied. The sessions all took place at moderate room density. There were 8 sessions with a sandwich marker, 13 with a sweater, and 20 with the books.

The authors were interested in whether a marker would reserve an entire table as well as the marked chair. The answer for all of the markers was affirmed. The unmarked control tables were occupied significantly sooner than were the marked tables, and the difference was significant for each of the three markers. In fact, in all 41 sessions the control table was occupied sooner or at the same time as the marked table. In only three of the sessions did anyone sit at the marked *chair*. All three were occupied by males, a finding whose significance will be discussed later. It is also interesting to examine the occupancy patterns at the two sorts of tables. The marked tables were eventually occupied by 34 lone individuals and 4 groups of 2 persons, while the unmarked tables were occupied by 18 lone individuals and 20 groups. It can be noted that a group of 2 or 3 could easily be accommodated at a marked table even assuming that the marker represented one person, yet virtually all the groups sat at unmarked tables. It is clear that the markers were able to (*a*) protect the particular chair almost totally, (*b*) delay occupancy of the entire table, and (*c*) divert groups away from the table.

STUDY 3

A similar study using books and newspapers as markers was undertaken in a dormitory study hall at a time of very light room density. Virtually all the markers proved effective in reserving the marked chair. The only exceptions were two sessions when the school paper which had been used as a marker was treated as litter and pushed aside. After more than 30 individual sessions where virtually all the markers were respected, the authors decided to move the experiments to the main university library where room density was much heavier. It seemed clear that at low densities almost any marker is effective. One qualification is that the object must be perceived as a marker and not as something discarded and unwanted by its former owner. Certain forms of litter such as old newspapers or magazines may, indeed, attract people to a given location.

The locus of study was switched to the periodical room in the university library where room density was high and pressure for seats was great. This room contained rectangular six-chair tables, three chairs to a side. The experimenter arrived at one of the six seats at a designated table at 6:50 P.M., deposited a marker, and then departed to another table at 7:00 P.M. to view any occupancy at the marked position by a student seeking space. During each session, a similarly situated empty chair which was unmarked was used as the control. There were 25 experimental sessions, each lasting 2 hours. The markers included two notebooks and a textbook, four library journals piled in a neat stack, four library journals randomly scattered on the table, a sports jacket draped over the chair, and a sports jacket draped over the chair in addition to the notebooks on the table.

If one compares the average time before occupancy of the marked and the control chairs, it is apparent that all markers were effective. Seventeen of the 25 marked chairs remained vacant the entire 2-hour period, while *all* control chairs were occupied. The average interval before the control chairs were occupied was 20 minutes. Some of the markers were more potent than others. Only one student occupied a chair that was marked either by a sports jacket or a notebook-and-text. Chairs marked by the neatly piled journals were occupied three of the five sessions, while chairs marked by the randomly placed journals were occupied all five sessions, even though the interval in each case exceeded that of the control chairs. It is clear that the personal markers, such as the sports jacket and notebooks, were able to keep away intruders entirely, while the impersonal library-owned markers (journals) could only delay occupancy of the marked chairs.

An interesting sidelight is that eight of the nine students who sat down despite the markers were males. Since there were more females than males in the control chairs at the same time, the high incidence of males is quite significant. It may be recalled in the previous study that the only three individuals who pushed aside the marker and sat at a marked chair were also males. It is likely that some sort of dominance or risk-taking factor is at work in the decision to disregard a territorial marker. The relationship between personality characteristics and the likelihood of invading someone else's space seems an exciting topic for further investigation.

Another serendipitous finding concerns the role of the neighbor, the person sitting alongside the marked chair, in defending the marked space. In all five trials with the scattered journals, the potential invader questioned the person sitting alongside the marked chair (the neighbor) as to whether the space was vacant. Early in the 2-hour session, the neighbor unknowingly served as the protector of the space. He informed all inquisitive intruders that the space was taken, since he believed the experimenter would return in view of the marker left on the table. As time passed, the neighbor's belief that the experimenter would return to the chair began to wane. At this point he would impart his new conception of the situation to potential invaders, "Yes, somebody was sitting there, but that was over an hour ago. Maybe he's not coming back."

STUDY 4

Since the role of the neighbor seemed an important aspect of a property-ownership system, the authors decided to investigate it experimentally. The first of such studies involved two experimenters and a person sitting alongside an empty chair. One experimenter seated himself next to a stranger (the neighbor) for 15 minutes and then departed, leaving behind an open book and an open notebook upon the table as territorial markers. After a fixed interval, the second experimenter, in the role of a student looking for a chair, came and inquired about the marked space nonverbally. The nonverbal questioning was a pantomime which included catching the neighbor's eye, pulling out the chair slightly, hesitating, looking at the place markers and at the neighbor, and then

back at the markers. The authors had very little experience with such nonverbal cues, but expected that the neighbor's reactions might include verbal defenses ("That seat is taken") and nonverbal defenses (moving the books to reinforce the marker). The independent variable was the length of time between the departure of the first experimenter and the arrival of the second—which was either a 5- or a 20-minute interval. Some sessions had to be terminated when the neighbor departed before the second experimenter arrived on the scene.

Overall the results were discouraging. in only 6 of the 55 trials did the neighbor respond to the nonverbal gestures of the second experimenter in what could be described as a space-defending manner, such as a statement that the seat was taken. Five of the six defensive acts occurred when the experimenter had been away 5 minutes, compared to only one defensive act when he had been away 20 minutes, but considering that there were 55 trials the difference was unimpressive.

STUDY 5

The authors decided to make another attempt to see if the neighbor could be involved in property defense on a spontaneous basis—that is, if he would defend marked space without being questioned directly. Unlike in the preceding study, the "owner" attempted to establish a relationship with the neighbor prior to the "owner's" departure. There were two phases of the study; when it seemed that the first approach was not leading anywhere, another approach was used. The markers were a neat stack of three paperback books left on the table in front of a chair. The sessions took place at six-chair tables where there was at least 1 empty seat between the marker and the neighbor. The first experimenter entered the room and found the location meeting the experimental requirements (a person sitting at the end chair of a six-person table with two empty chairs alongside him—O-O-S). The experimenter (a girl) sat down on the same side of the table but one seat away (E-O-S). There were 13 trials in each of the following conditions: (*a*) The experimenter sat 5 minutes and then departed from the table, leaving her books neatly stacked on the table. During this time she did not interact with her neighbor. (*b*) Similar to Condition *a*, the experimenter sat for 5 minutes except that during the 5-minute wait, the experimenter asked the neighbor "Excuse me, could you tell me what time it is?" (*c*) Similar to Condition *a*, the experimenter sat for 5 minutes except that during the 5-minute wait the experimenter engaged the neighbor in conversation four times and, while leaving and placing the stack of three paperback books on the table, declared, "See you later." Fifteen minutes later, the second experimenter (a male) entered the room, walked directly to the marked chair, pushed the books directly ahead of him, and sat down at the table.

The results were again discouraging. In none of the 39 trials involving Conditions *a, b,* and *c* did the neighbor inform the intruder that the seat was taken. The authors therefore decided to strengthen the conditions by having the "owner" return and directly confront the intruder. Seven of such trials were added to Condition *a*, six to Condition *b*, and 6 to

Condition *c*, making 19 trials in all when the "owner" came back and told the intruder "You are sitting in my chair." Each time she hesitated about 30 seconds to see if the neighbor would intervene, and then she picked up her books and departed. There was no verbal response from the neighbor in any of the 19 sessions. The most that occurred would be a frown or a look of surprise on the part of the neighbor, or some nonverbal communication with someone else at the table. Stated simply, despite a flagrant usurpation of a marked space, all neighbors chose to remain uninvolved. It became clear that if one wanted to study the neighbor's role in such an informal regulatory system one would have to question him directly as to whether the seat was occupied.

STUDY 6

The next study employed two experimenters, a male and a female, and the same three paperback books as markers. Two different girls were used as experimenters, and the sessions occurred in two different, nearby college libraries. The experimental situation involved six-chair tables where the first experimenter (female) sat down at the same side of a table with a subject, leaving an empty chair between them (E-O-S). The goal of the study was to learn whether a greater amount of interaction between the former occupant and the neighbor would increase the neighbor's likelihood of defending the chair. Unlike in the previous study, the neighbor was questioned directly as to whether the seat was taken. There were three different instructional sets, and these took place according to a prearranged random order. In 14 trials, the first experimenter sat at the chair for 5 minutes without saying anything, deposited the marker (three paperback books), and left. Fourteen other sessions were similar except that at some time during her 5-minute stay, the first experimenter asked the neighbor for the time. Ten other sessions were similar except that the experimenter engaged the neighbor in conversation as to where to get a coke, what was happening on campus, and other minor matters. Fifteen minutes after the first experimenter departed, the second experimenter (a male) entered the room, walked over to the marked chair, and asked the neighbor "Excuse me, is there anyone sitting here?"

The results differ markedly from those in the previous study. A total of 22 out of the 38 neighbors defended the seat when questioned directly on the matter. The typical defense response was "Yes, there is" or "There is a girl who left those books."[3] However, the amount of contact between the first experimenter and the neighbor made little difference in defensive behavior. When there had been no contact, or minimal contact, between the first experimenter and neighbor the seat was protected 58% of the time, while the use of several items of conversation between the experimenter and her neighbor raised the percentage of defensive responses

[3]The neighbors' replies to the intruder's question were scored separately by two coders as indicating defense of the space ("Yes, that seat is taken") or nondefense ("No, it isn't taken" or "I don't know"). There was 100% agreement between the two raters in scoring the replies into defense or nondefense categories.

only to 66%. The difference between conditions is small and statistically unreliable; what is impressive is the great increase in defensive behavior when the neighbor was questioned directly. Two other parameters of the situation are (*a*) the time that the first experimenter remained in the seat before depositing her marker, and (*b*) the length of time that the first experimenter was out of room before the second experimenter approached the marked chair.

STUDY 7

The final study employed two experimenters, both males, and the same three paperback books. The sessions took place at six-chair tables in the library, where the first experimenter again sat down on the same side of the table with a subject, leaving an empty chair between them (E-O-S). He remained either 5 minutes or 20 minutes, depending upon the experimental condition, and then departed, leaving on the table a neat stack of three paperback books. After a designated interval of either 15 or 60 minutes, the second experimenter entered the room and asked the neighbor whether the (marked) chair was taken. The second experimenter recorded the neighbor's reply verbatim just as soon as he was able to sit down somewhere. Since both experimenters were male, it was decided to use only male neighbors in the experiment.

The independent variables were (*a*) the length of time the first experimenter had been seated before he left his marker and departed and (*b*) the length of time the first experimenter was absent before the neighbor was questioned by the second experimenter. Some sessions were unusable since the neighbor departed before the designated time and could not be interviewed. Most of the unusable sessions occurred when the experimenter had been absent for 60 minutes. The sessions took place at times of light-to-moderate room density.

Although the design had not called for comparison of marked and unmarked chairs, it is noteworthy that the markers were effective in keeping people away. Not one of the 64 marked chairs was ever occupied. Regarding the inclination of the neighbor to defend the marked space when questioned by the second experimenter, a content analysis of the neighbor's responses to the query "Is this seat taken?" into defense and nondefense categories revealed that 44 neighbors defended the marked space by indicating that it was taken, while 20 failed to do so either by pleading ignorance or by stating that the chair was empty. The response to a direct question stands in contrast to the lack of involvement when neighbors were approached nonverbally. The length of time that the first experimenter had originally occupied the chair (his tenure period) had no effect on the willingness of the neighbor to defend the chair. However, the length of time that the previous owner was away—either 15 or 60 minutes—had a significant effect. When the former owner had been absent 15 minutes, 80% of the neighbors defended the space compared to 54% defending it when the former owner had been away a full hour ($p < .05$).

Several aspects of the results require

elaboration. It is possible that initial tenure periods of 5 and 20 minutes were not sufficiently different. Yet it seems noteworthy that even with a rather impersonal marker, more than two-thirds of the neighbors defended the marked chair upon direct questioning. Most of those who didn't defend it simply pleaded ignorance ("I don't know if it's taken") rather than indicating that the seat was vacant.

After the experiments had been completed, 15 additional students in the library were interviewed on the question of how personal belongings could reserve space. Each student was asked how he would react if he saw someone intrude into a marked space, particularly if the original owner came back and claimed the space (i.e., the actual experimental situation was described to him). The replies were at variance with what the authors had actually found in such a situation. Most of the respondents maintained that they would indeed protect a marked space, although some of them added qualifications that they would defend the space only if the person were away a short time. Typical responses were: "I would protect the person's books and state (to the intruder) that the place was obviously taken by the presence of the books," and "Yes, I would mention that someone was sitting there." Although the majority mentioned specifically that they would protect a marked chair, in the actual situation no one had done so unless approached directly. The ethic regarding space ownership in the library exists, but is paid lip service, probably because institutional means of enforcement do not exist.

Discussion

The present article represents a small beginning toward understanding how markers reserve space and receive their legitimacy from people in the area (neighbors) and potential intruders. Psychologists have paid little attention to boundary markers in social interaction, perhaps because such markers were regarded as physical objects relegated to the cultural system (the province of the anthropologist) rather than an interpersonal system which is the true province of the social psychologist. Generally it is the geographers and lawyers who are most concerned with boundaries and markers. Since the present studies took place in public spaces, we are dealing more with norms and customs than with legal statutes. Stated another way, the situations involve an interpersonal system where sanctions are enforced by the individuals immediately present. Goffman (1963) labels the situations the authors used in the experiments *temporary territories*. It is clear that a person placing his coat over the back of a chair desires to reserve the space, and most people in the immediate vicinity will support his claim if questioned (although they will remain uninvolved if they can); such behavior meets Hediger's (1950) definition of territory presented previously as well as the more simple one provided by Noble (1939) that a territory represents "any defended area." The phenomena the present authors have studied do not belong under other available rubrics of spatial behavior, such as home range, biotope, niche, or life space. The major differences between the

experimental situations and more enduring territories is that the latter are meshed with a legal-cultural framework and supported in the end by laws, police, and armies. The marked spaces in the present authors' experiments have no legal status and are supported only by the immediate social system. Occasionally it became necessary to articulate the structure of the system by "requiring" neighbors to enter the situation.

People are now spending an increasing portion of their time in public or institutional spaces, including theaters, airport lobbies, buses, schools, and hospitals, where the use of personal belongings to mark out temporary territories is a common phenomenon. The study of territories, temporary as well as enduring ones, deserves study by psychologists. There is some danger that such work will lose much of its force if some semantic clarity is not obtained. While the ethologist's definition of a territory as "any defended area" has considerable heuristic value, there is no need to assume that the mechanisms underlying human and animal behavior are identical. The paucity of data about human territorial behavior makes it most reasonable to assume that the mechanisms are analogous rather than homologous.

In conclusion, the present series of studies suggests that further investigation of spatial markers is feasible and warranted. The physical environment has for too long been considered the background variable in psychological research. The time is past when we can have theories of man that do not take into account his surroundings. Boundary markers not only define what belongs to a person and what belongs to his neighbor, but also who he is and what it means to be a neighbor in a complex society.

References

Altman, I., & Haythorn, W. W. The ecology of isolated groups. *Behavioral Science*, 1967, **12**, 169–182.

Ardrey, R. *African genesis*. London: Collins, 1961.

Ardrey, R. *The territorial imperative*. New York: Atheneum, 1966.

Esser, A. H. et al. Territoriality of patients on a research ward. In J. Wortis (Ed.), *Recent advances in biological psychiatry*. Vol. 8. New York: Plenum Press, 1965.

Felipe, N., & Sommer, R. Invasions of personal space. *Social Problems,* 1966, **14**, 206–214.

Goffman, E. *Behavior in public places*. New York: Free Press of Glencoe, 1963.

Hall, E. T. *The hidden dimension*. Garden City: Doubleday, 1966.

Hediger, H. *Wild animals in captivity*. London: Butterworths, 1950.

Lipman, A. Old peoples homes: Siting and neighborhood integration. *The Sociological Review*, 1967, **15**, 323–338.

Lyman, S. M., & Scott, M. B. Territoriality: A neglected sociological dimension. *Social Problems*, 1967, **15**, 236–249.

Noble, G. K. The role of dominance in the social life of birds. *Auk*, 1939, 263–273.

Roos, P. D. Jurisdiction: An ecological concept. *Human Relations*, 1968, **21**, 75–84.

Sommer, R. Sociofugal space. *American Journal of Sociology*, 1967, **72**, 654–660.

Exercise

Witnessing Behavior of Others: The Role of Models in Affecting the Audience

Imitation plays an important role in learning new behaviors and performing old behaviors. Adults as well as children are likely to perform acts which they have observed committed by others, whether the others appear in person or on a television screen (Bandura, 1969; Bryan and Schwartz, 1970). Among the kinds of behavior that have been shown to be affected by the observations of others are aggressive responses, self-sacrificing acts, and language habits. For example, subjects are more likely to hit a Bobo doll and to donate money to charity if they have witnessed a model perform one of these acts.

Models may elicit matching behaviors for a variety of reasons and there is considerable controversy among theorists concerning the exact mechanisms that govern observational learning and imitative performance. Reinforcement for a model's behavior may create a generalized habit of imitation, that is, if a child has been reinforced for twenty specific acts of imitation, he may develop a tendency to imitate many acts. In situations where appropriate behavior is not clearly defined, a model may simply serve to suggest a proper form of action; the model's behavior serves as a guide for the observer. Although the processes that mediate such imitation are not completely understood, the phenomenon, imitation, is extremely important in governing human behavior.

This exercise, which is based on Experiments II and III in the Research Example by Bryan and Test, is designed to demonstrate modeling. The actions of the model as he donates money to a charity will serve as the independent variable. The dependent variable will be the observer's behavior. Thus, the experiment is concerned with the impact of a model's charitable

act on the subsequent behavior of his audience, who in this case are passersby. The student may find it entertaining and helpful to think of alternative tasks and situations that better fit his particular interests and settings. For example, one might study the impact on observers either of a charity donor or of someone in a supermarket who gives his place in line to someone with fewer items than himself.

METHOD

Experimenters

Teams of two students are required for the data collection. One student will serve as the model, the other as the data recorder. The model's task will be to influence the behavior of fellow pedestrians with his own actions.

Setting

Sufficient pedestrians should pass through the test area so that long waits between subjects will not be necessary. On the other hand, traffic should not be so heavy that the recorder cannot clearly identify the subject. Furthermore, the location should minimize the obtrusiveness of the recorder and the model so that they will not seem "out of place." In an ideal setting the recorder could remain hidden during his observations. A convenient building or parked car can serve as an observation post and as a resting place for the model. A good setting will minimize the visibility of the model to pedestrians other than the test subject. For example, you might conduct the experiment on a street corner so that only pedestrians who have already rounded the corner are included in the study. Figure 1 shows such an arrangement.

In Figure 1, only subjects coming from the north and turning east would be included in the study. This setting minimizes the number of potential subjects who must be eliminated because they have observed a previous subject being run. This type of experiment can be done inside a building or outdoors.

Apparatus

You will need enough props to create a plausible and legitimate donation context. Donation cannisters may easily be obtained from the March of Dimes; other reputable charities may be willing to loan posters and collection boxes. Campus organizations conducting fund-raising drives might also help. Obviously the charity or fund drive will benefit from the money collected by your experimental team.

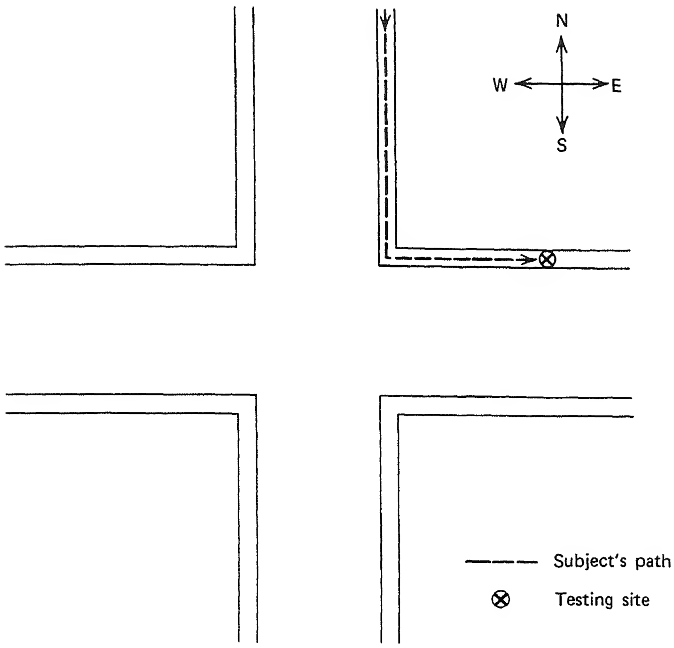


Figure 1. Street corner testing site.

DATA COLLECTION; INDEXING IMITATIVE ALTRUISM

Because of the difficulties involved, we shall first discuss the problems of indexing or measuring imitation before detailing the other aspects of the procedure. To measure the impact of a model, you will compare two groups of subjects: those who are exposed to a donating model versus control subjects who do not observe a model. The comparison of these two groups is achieved by calculating the percentage of donators within each group and computing a chi-square analysis (See Appendix B).

Please Note This Precaution.

The chi-square test requires independent observations. That is, it requires that individuals be *independently* exposed to the treatment conditions (model versus no model situations). The behavior of one subject cannot be allowed to influence another subject. If this intersubject influence occurs, these observations cannot be included in a chi-square analysis. Thus you cannot allow the members of a group to influence one another. In the light of this restriction, how should you score the behavior of groups of people who happen to see the model at the same time? For example, suppose a group of three persons approach the donation cannister, witness the model donate, and then proceed to donate. Would you indicate that three people or one person imitated the model?

Two procedures may be used to obtain the conditions of independence required. First, you may wish to use only pedestrians who approach the cannister alone and have no opportunity to be affected by members of a group. Thus, unaccompanied individuals as well as those who observe others' actions would not be included in the tally. What limitations would this place on generalizing from your results? For example, such a sample might be dominated by "loners" who are seldom with other people. Would such people be more or less likely to be influenced by social models?

A second and usually more convenient method of obtaining independent subjects is to treat groups of people as you would an individual person in tallying your results. Thus, if any member of a group donated, the tally of the number of affected persons would be one, not the actual number of people giving money. For example, if all three members of a group contributed to the charity, you would treat this observation as a single act—that is, you would score the group action the same as the action of a single person by assigning it a score of one. Similarly, if all members of a ten member group contributed, you would tally this by indicating that one person, not ten, emitted that response; if three out of ten contribute, the score would be one, not three. For *each occasion* only one of two scores is possible, a one or a zero. These scores should be recorded on the Data Sheet provided on page 209.

Procedure

Before testing subjects, make a comprehensive list of the criteria you will use to select subjects. Do you intend to test only females, males, unaccompanied pedestrians, or groups of pedestrians? Next, decide those conditions that define the initiation of a new test trial. For example, after the subjects who have been exposed to the model have been tested, when will you begin the next no-model treatment? Will you start the next trial after a certain period of time has elapsed (for example, 30 seconds) or after a certain number of pedestrians have passed? Some of these problems are discussed in Exercise 1.

You must adhere carefully to your chosen procedures until the completion of the experiment. If you find that your decisions are not practical, redesign the experiment and start again. Once you have found convenient procedures, data collection should be rapid and the experiment can be completed within a few hours.

For the altruistic model condition, have the model approach the donation cannister, contribute a small amount of money, and leave. Make sure that the subject can easily observe the model's actions. You might find it beneficial to add a third number to your research team. Having this third

person stand by the cannister, apparently attempting to solicit funds, may heighten the model's influence on the observer. If you choose to add a third member, make sure that this solicitor is visually and verbally salient in the same degree to all subjects. If he is more visible to one group than to another, he might unwittingly bias your findings (See discussion of bias in Exercise 1.)

For the no-model condition, either have the model walk by the cannister without donating (selfish model) or use no model at all. If you decide upon the latter, the treatment condition lasts until the subject has passed the cannister. That is, the duration of the treatment condition ends when it is reasonable to assume that the subject has decided whether to donate to the charity.

Alternating the modeling manipulations. First test one subject in a certain condition; follow this by testing a subject in the other treatment condition. Since modeling effects are likely to be weak in such a setting, you should test many subjects in order to obtain a significant effect. Since testing subjects is easy and fast in a good location, you should be able to test 50 to 100 subjects within each condition in less than a day.

Data Analysis

The appropriate test for the hypothesis that the observation of altruistic models increases the altruistic behavior of witnesses is a chi-square test of significance (see Appendix B) performed on the data in the Frequency Table at the bottom of the Data Sheet, page 209.

References

Bandura, A. *Principles of behavior modification*. New York: Holt, Rinehart & Winston, 1969.

The most complete presentation of the literature on modeling and its therapeutic applications.

Bryan, J. H., and Schwartz, T. Effects of film material upon behavior. *Psychological Bulletin*, 1970, **75**, 50–59.

This article reviews film effects on behavior. Much of the article is devoted to descriptions of imitative behavior of televised models.

Lefkowitz, M., Blake, R. R., and Mouton, J. S. Status factor in pedestrian violation of traffic signals. *Journal of Abnormal and Social Psychology*, 1955, **51**, 704–706.

An interesting and ingenious study concerning the role of status (as indexed by wearing apparel) in imitative behavior. This is particularly interesting as it was conducted in a naturalistic setting and not a laboratory.

Wagner, C., and Wheeler, L. Model, need, and cost effects in helping behavior. *Journal of Personality and Social Psychology*, 1969, **12**, 111–116.

An interesting laboratory experiment on the role of altruistic models in evoking high-cost behavior from adult witnesses.

Data Sheet 1

Experimental Group
Subjects Exposed to
Altruistic Model

S No.	Donated	
	Yes	No
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13*		
TOTAL	A	B

Comparison Group
Subjects Not Exposed to
Altruistic Model

S No.	Donated	
	Yes	No
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13*		
TOTAL	C	D

FREQUENCY TABLE

Donated

Group:	Yes	No
Altruistic Model	*A*	*B*
Comparison	*C*	*D*

To compute chi-square, transfer numbers to appropriate boxes in Appendix B.

*If more subjects are used, continue numbering and record data on a similar page. Carry totals to the Frequency Table.

Research Example

Models and Helping: Naturalistic Studies in Aiding Behavior[1]

JAMES H. BRYAN AND MARY ANN TEST

4 experiments concerned with helping behavior were conducted. 3 were addressed to the effects of altruistic models upon helping, while 1 was concerned with the impact of the solicitor's race upon donations. 3 investigations employed as a site parking lots of 2 large department stores in New Jersey, and indexed helping by contributions to the Salvation Army. A 4th experiment indexed helping by offers of aid by passing motorists to a woman with a disabled vehicle. Whether one employed motorists in California or shoppers in New Jersey, the results were quite consistent. The presence of a helping model significantly increased helping behavior. As race of the Salvation Army solicitor did affect the percentage of donors willing to contribute money, it was concluded that interpersonal attraction is a relevant variable affecting donations.

Recently, concern has been evidenced regarding the determinants and correlates of altruistic behavior, those acts wherein individuals share or sacrifice a presumed positive reinforcer for no apparent social or material gain. Studies addressed to these behaviors have explored both individual differences in the tendency to be altruistic and the situational determinants of such responses. Gore and Rotter (1963) found that students at a southern Negro college were more likely to volunteer for a social protest

SOURCE: *Journal of Personality and Social Psychology,* 1967, **6**, (4), 400–407.

[1] While Mary Ann Test collaborated with the senior author on Experiment I, the remaining work is the latter's sole responsibility.

Thanks are due to Cheryi Dellhoussay, Betty Umann, Joe McNair, and Frank Siri who served as the experimenters and stooges for Experiment I, and to Edward Nystrom, Alice Anderson, Katherine Moore, and Irene Paramoure who served as the models, observers, and solicitors in studies II, III, and IV. Studies II, III, and IV were carried out while the author was affiliated with Educational Testing Service and were supported by the National Institute of Child Health and Human Development, under Research Grant 1 PO1 HD1762-01. The authors are especially grateful to the Salvation Army of Trenton, New Jersey, and specifically to George H. Gibb, whose cooperation made these experiments possible. Thanks are also due to Perry London, David Rosenhan, Ladd Wheeler, Lawrence Stricker, and Bruce K. Eckland for the many helpful comments upon various portions of the manuscript.

movement if they perceived sources of reinforcement as internally rather than externally guided. Subjects high on internal control were more likely to volunteer as freedom riders, marchers, or petition signers than subjects who perceived others as primary agents of reinforcement. Experimental evidence has been generated supporting the often-made assumption that guilt may serve as a stimulus to altruistic activity. Darlington and Macker (1966) found that subjects led to believe that they had harmed another through incompetent performances on the experimental tasks (three paper-and-pencil tests) were more willing than control subjects to donate blood to a local hospital. Aronfreed and Paskal[2] and Midlarsky and Bryan (1967) found that children exposed to treatment conditions designed to produce empathy were more willing to donate M&M candies than subjects given control conditions, while Handlon and Gross (1959), Ugurel-Semin (1952), Wright (1942), and Midlarsky and Bryan have found sharing to be positively correlated with age among school-age children. Lastly, Berkowitz and Friedman (1967) have demonstrated that adolescents of the working class and the bureaucratic middle class are less affected in their helping behaviors by interpersonal attraction than adolescents of the entrepreneur middle class.

Three hypotheses have emerged regarding the situational determinants of self-sacrificing behaviors. One suggests that individuals behave in an altruistic fashion of compliance to a norm of reciprocity. That is, individuals are aware of the social debts and credits established between them, and expect that ultimately the mutual exchange of goods and services will balance (Gouldner, 1960). Berkowitz and Daniels (1964) have suggested that individuals might show a generalization of such obligatory feelings and thus aid others who had not previously assisted them.

A second hypothesis was put forth by Berkowitz and his colleagues (Berkowitz, 1966; Berkowitz & Daniels, 1963; Berkowitz, Klanderman, & Harris, 1964; Daniels & Berkowitz, 1963) who have postulated the social responsibility norm. They have contended that dependency on others evokes helping responses even under conditions where the possibility of external rewards for the helper are remote. Using supervisor's ratings of an unknown and absent other to produce dependency, and a box-construction task as the dependent variable, considerable support has been generated for the suggestion that dependency increases helping.

A third major determinant of helping may be the presence of helping (or nonhelping) models. While attention to the effects of models has generally been directed toward antisocial behaviors (cf. Bandura & Walters, 1963; Freed, Chandler, Mouton, & Blake, 1955; Lefkowitz, Blake, & Mouton, 1955), some recent evidence suggests that observation of self-sacrificing models may lead to subsequent succorant behavior by children. For example, Rosenhan and White (1967) have demonstrated that children

[2] J. Aronfreed & V. Paskal. Altruism, empathy and the conditioning of positive affect. Unpublished manuscript, 1965.

are more likely to donate highly valued gift certificates to residents of a fictitious orphanage if they have seen an adult do so. Hartup and Coates[3] found that nursery school children who have been exposed to a self-sacrificing peer were more likely to be altruistic than children not so exposed. Test and Bryan[4] found that female college students were more likely to render aid to another in computing arithmetic problems if they saw other people so doing.

The present series of experiments was designed to test the effects of models in natural settings on subject samples other than college or high school students, and in context other than a school room or university setting. The first three experiments reported are concerned with the impact of observing helping models upon subsequent helping behaviors, while the fourth is addressed to the influence of interpersonal attraction upon donation behavior.

Experiment I: Lady in Distress: A Flat Tire Study

Few studies have been concerned with the effects of models upon *adults*, and fewer still with the impact of *prosocial* models upon them (Wheeler, 1966). Those that have been concerned with such behaviors have invariably employed college students as subjects. For example, Rosenbaum and Blake (1955) and Rosenbaum (1956) have found that college students exposed to a model who volunteered, upon the personal request of the experimenter, to participate in an experiment would be more likely to consent than subjects not exposed to such a model or than subjects who observed a model refuse to cooperate. Pressures toward conformity in these experiments were great, however, as the request was made directly by the experimenter and in the presence of a large number of other students.

Test and Bryan found that the observation of helping models significantly increased the subsequent offers of aid by observers. However, in that study, subjects were given the task of solving arithmetic problems and then rating their difficulty, a task ordinarily requiring autonomous efforts. Furthermore, the experiment was conducted within a university setting, a context where independence of thought is often stressed. The effects of the model may have been simply to increase the subjects' faith that assisting others was allowed. While questionnaire data of the study did not support this interpretation, such effects could not be ruled out entirely. Thus, it is possible that the model impact was simply a propriety-defining activity which reduced the inhibitions associated with such helping behavior.

In general, then, investigations of modeling that employ adults as subjects and that demand self-sacrifice on the part of subjects are limited in number, exploit strong pressures toward conform-

[3] W. W. Hartup & B. Coates. Imitation of peers as a function of reinforcement from the peer group and rewardingness of the model. Unpublished manuscript, 1966.

[4] M. A. Test & J. H. Bryan. Dependency, models and reciprocity. Unpublished manuscript, 1966.

ity, and rely upon college students as subjects. The present experiment was designed to assess the impact of models upon subsequent spontaneous offers of help in other than a university setting.

METHOD

The standard condition consisted of an undergraduate female stationed by a 1964 Ford Mustang (control car) with a flat left-rear tire. An inflated tire was leaned upon the left side of the auto. The girl, the flat tire, and the inflated tire were conspicuous to the passsing traffic.

In the model condition, a 1965 Oldsmobile was located approximately ¼ mile from the control car. The car was raised by jack under the left rear bumper, and a girl was watching a male changing the flat tire.

In the no-model condition, the model was absent; thus, only the control car was visible to the passing traffic.

The cars were located in a predominantly residential section in Los Angeles, California. They were placed in such a manner that no intersection separated the model from the control car. No turnoffs were thus available to the passing traffic. Further, opposite flows of traffic were divided by a separator such that the first U turn available to the traffic going in the opposite direction of the control car would be after exposure to the model condition.

The experiment was conducted on two successive Saturdays between the hours of 1:45 and 5:50 P.M. Each treatment condition lasted for the time required for 1000 vehicles to pass the control car. While private automobiles and trucks, motorscooters, and motorcycles were tallied as vehicles, commercial trucks, taxis, and buses were not. Vehicle count was made by a fourth member of the experiment who stood approximately 100 feet from the control car hidden from the passing motorists. On the first Saturday, the model condition was run first and lasted from 1:45 to 3:15 P.M. In order to exploit changing traffic patterns and to keep the time intervals equal across treatment conditions, the control car was moved several blocks and placed on the opposite side of the street for the no-model condition. The time of the no-model treatment was 4:00 to 5:00 P.M. On the following Saturday, counterbalancing the order and the location of treatment conditions was accomplished. That is, the no-model condition was run initially and the control car was placed in the same location that it had been placed on the previous Saturday during the model condition. The time of the no-model condition was 2:00 to 3:30 P.M. For the model condition, the control car was placed in that locale where it had been previously during the no-model condition. The time of the model condition was 4:30 to 5:30 P.M.

Individuals who had stopped to offer help were told by the young lady that she had already phoned an auto club and that help was imminent. Those who nonetheless insisted on helping her were told the nature of the experiment.

RESULTS

The dependent variable was the number of cars that stopped and from which at least one individual offered help to

the stooge by the control car. Of the 4000 passing vehicles, 93 stopped. With the model car absent, 35 vehicles stopped; with the model present 58 halted. The difference between the conditions was statistically significant ($\chi^2 = 5.53$, corrected for continuity, $df = 1$, $p < .02$,[5] two-tailed). Virtually all offers of aid were from men rather than women drivers.

The time of day had little impact upon the offering of aid. Fifty vehicles stopped during the early part of the afternoon; 43 during the later hours. Likewise, differences in help offers were not great between successive Saturdays, as 45 offers of aid were made on the first Saturday, 48 on the second Saturday.

The results of the present study support the hypothesis that helping behaviors can be significantly increased through the observation of others' helpfulness. However, other plausible hypotheses exist which may account for the findings. It is possible to account for the differences in treatment effects by differences in sympathy arousal. That is, in the model condition, the motorist observed a woman who had had some difficulty. Such observations may have elicited sympathy and may have served as a reminder to the driver of his own social responsibilities.

Another explanation of the findings revolves around traffic slowdown. It is possible that the imposition of the model condition served to reduce traffic speed, thus making subsequent stopping to help a less hazardous undertaking. While the time taken for 1000 autos to pass the control car was virtually identical in the model and no-model condition and thus not supportive of such an explanation, the "slowdown" hypothesis cannot be eliminated. Assuming the model effect to be real, one might still argue that it was not a norm of helping that was facilitated by the model, but rather that inhibitions against picking up helpless young ladies were reduced. That is, within the model condition, the passing motorists may have observed a tempted other and thus felt less constrained themselves regarding similar efforts. Indeed, the insistence of some people to help in spite of the imminent arrival of other aiders suggested the operation of motives other than simply helping. Indeed, while the authors did not index the frequency of pick-up attempts, it was clear that a rather large number were evidenced.

Because of the number of alternative explanations, the evidence supporting the hypothesis that the observation of helpers per se will increase subsequent aiding is weak. Experiment II was designed to test further the prediction that the perception of another's altruistic activity would elicit similar behavior on the part of the observer.

Experiment II: Coins in the Kettle

The investigation was conducted on December 14th between the hours of 10:00 A.M. and 5:00 P.M. The subjects were shoppers at a large department store in Princeton, New Jersey. Observations made on the previous day indicated that the shoppers were overwhelmingly Caucasian females.

[5] Notations such as $p < .02$ are explained in the Introduction, pp. 15-16.

A Salvation Army kettle was placed on the sidewalk in front of the main entrance to the store. Two females, both in the experimenter's employ, alternately manned the kettle for periods of 25 minutes. One solicitor was a Negro, the other a Caucasian. Each wore a Salvation Army cape and hat. Although allowed to ring the Salvation Army bell, they were not permitted to make any verbal plea or to maintain eye contact with the passing shoppers, except to thank any contributor for his donation.

The model condition (M) was produced as follows: Once every minute on the minute, a male dressed as a white-collar worker would approach the kettle from within the store and contribute 5 cents. As the model donated, he started a stopwatch and walked from the kettle toward a parking lot as if searching for someone. He then returned to the store. The following 20-second period constituted the duration of the treatment condition.

Following a subsequent lapse of 20 seconds, the next 20-second period defined the no-model condition (NM). Within any one minute, therefore, both M and NM treatments occurred. There were 365 occasions of each treatment.

It should be noted that it was possible that some subjects in the NM condition observed the contribution of the model or a donor affected by the model. If that hypothesis is correct, however, the effects of such incidents would be to reduce rather than enhance the differences between treatments.

RESULTS

The dependent variable was the number of people who independently donated to the Salvation Army. People obviously acquainted, as for example, man and wife, were construed as one potential donating unit. In such conditions, if both members of a couple contributed, they were counted as a single donor.

Since there were no differences in model effects for the Negro or Caucasian solicitor, data obtained from each were combined. The total number of contributors under the NM condition was 43; under the M condition, 69. Assuming that the chance distribution of donations would be equal across the two conditions, a chi-square analysis was performed. The chi-square equaled 6.01 ($p < .01$).[6]

In spite of precautions concerning the elimination of correlated observations within a treatment condition, it was possible for subjects in any one observational period to influence one another. Such influence may have been mediated through acquaintances not eliminated by our procedures or the observations of others as well as the model donating. A more conservative analysis of the data, insuring independent observation, was therefore made. Instead of comparing treatments by analyzing the number of donors, the analysis used, as the dependent variable, the number of observation periods in which there was a contribution, that is, those periods in which more than one donation occurred were scored identically to those in which only a single contribution was received. Occasions of donations equaled 60 in the M treatment, 43 in the NM condition. The chi-square equaled 2.89 ($p < .05$).

[6] All chi-square analyses were corrected for continuity and all tests of significance were one-tailed.

The results of Experiment II further support the hypothesis that observation of altruistic activity will increase such behavior among observers. But the matter is not yet entirely clear, for when the observer saw the model donate he saw two things: first, the actual donation, and second, the polite and potentially reinforcing interaction that occurred between the donor and solicitor. Conceivably, the observation of an altruistic model, per se, who was not socially reinforced for his behavior, would have little or no effect on an observer. The third experiment was designed to examine this possiblity.

Experiment III: Coins in the Kettle II

The experiment was conducted at a Trenton, New Jersey, shopping center from the hours of 10:00 A.M. to 5:00 P.M. Again, the majority of the patrons were Caucasian females. It is likely, however, that these shoppers were of a lower socioeconomic status than those in the Princeton group.

Salvation Army kettles were placed before the main entrance of a large department store (Kettle 1) and a large food center (Kettle 2). The kettles were separated by more than 200 yards. During the first 120 observations (10:00 A.M. to 12:00 P.M.), two male college students, employed by the Salvation Army and wearing its uniform, manned the kettles. The site of the experiment was Kettle 1, except on those occasions where the worker took his "coffee break." At those times, data collection was centered at Kettle 2. An equal number of M and NM conditions were run at each site, although approximately two-thirds of the observational time was spent at Kettle 1. During the remaining 240 observational periods (1:00 P.M. to 5:00 P.M.) the same male worker and his spouse alternately manned Kettle 1. The wife was stationed by the kettle for 136 minutes, the male for 104 minutes. The experiment was conducted only at Kettle 1 during the afternoon period.

Solicitors were told to make no verbal appeals for donations or responses to the model upon his contribution. While they were not informed of the hypothesis underlying the experiment, they may well have deduced it. The model was the same as in Experiment II, and again was dressed as a white-collar worker.

The imposition of the treatment conditions were identical to those described in Experiment II with the following exceptions. Since the kettle was more visible at this site than at the previous one, 30-second rather than 20-second periods were used for each treatment. To simplify the procedures, no waiting periods between treatments occurred. Additionally, after donating, the model would return to the parking lot. There were a total of 360 occasions of each of the M and NM conditions.

RESULTS

The criteria defining a donor were identical to those outlined in Experiment I. Under the M condition, 84 donors were tallied; under the NM treatment, 56. The chi-square value was 4.86 ($p < .025$).

Since it was possible that one donor might have seen a donor other than the model receive social approval from the solicitor, the more conservative comparison of the treatments as outlined in

Experiment II was made. That is, treatments were compared by noting the number of observational periods in which any donation occurred. Therefore, those donors who may have been influenced by a contributor receiving the solicitor's thanks were excluded. Of the 360 observational periods under the M condition, there were 75 in which some donation was made. Of the 360 periods, 51 were marked by contributions. Chi-square yielded a value of 5.09 ($p <$ .025).

Experiment IV: Ethnocentrism and Donation Behavior

While Experiment III was conducted to eliminate the solicitor's explicit social approval as a mechanism underlying donation behavior, it is possible that the model's impact was due to the information communicated to the observer regarding the consequence of donations. Work by Bandura, Ross, and Ross (1963), for example, found that children observing a model rewarded for aggression would be more aggressive than children who had observed a model being punished for such behavior. Additionally, considerable data have been gathered within the university laboratory suggesting that interpersonal attraction may greatly influence the helping response. Berkowitz and Friedman (1967), Daniels and Berkowitz (1963), and Goranson and Berkowitz (1966) have suggested that positive affect increases the probability of low payoff helping behavior.

The present experiment was designed to assess the impact of the solicitor's race upon the donation behavior of shoppers. It was assumed that a Negro solicitor would be held in less esteem by Caucasian shoppers than a solicitor of their same race, and that such attitudes would affect contributions. While the applicability of the "consequence to the model" hypothesis in accounting for the model's effect was not tested directly, the study assesses the importance of interpersonal attraction in eliciting charitable behavior.

METHOD

The experiment was conducted on December 2 and 3 between the hours of 10 A.M. and 6 P.M. at the Trenton area site. The subjects were Caucasian shoppers at a large department store.[7] Three thousand seven hundred and three shoppers were observed; 2,154 females and 1,549 males. In order to reduce the possibility of including the same subject in the experiment on more than one occasion, tallies were made only of exiting shoppers.

Two Salvation Army kettles were placed at two store exits, their location being separated by approximately 75 yards. Two female solicitors, a Negro and a Caucasian, manned the kettles. Both were in their early twenties, wore the uniform of the Salvation Army, and were in the employ of the experimenter. Each was instructed to make no verbal appeals for donations and to avoid eye contact with the shoppers. After a period of 25 minutes, the girls rotated kettle assignments, and during the last

[7] As there were very few Negro donors ($N = 7$), analysis was confined to the behavior of Caucasian shoppers.

Table 1
Analysis of Variance of Percentage Donor Scores

	df	MS	F
Race (A)	1	38.778	4.84[a]
Day (B)	1	98.315	12.28[b]
Kettle (C)	1	.018	
AXB	1	1.511	
AXC	1	11.340	
BXC	1	1.031	
AXBXC	1	3.206	
Error	48	8.009	

[a] $p < .05$ (2-tailed).

[b] $p < .01$ (2-tailed).

10 minutes of the hour were allowed to take a coffee break. Hence, during a single hour, each solicitor manned both kettles. Each solicitor manned each kettle on seven occasions per day. Thus, each solicitor was observed for a total of 28 observational periods; 14 on each day (seven on each kettle) over a period of two days.

Two observers, each assigned to a particular kettle, tallied the number and sex of the exiting shoppers and contributors during each of the 25 minute periods. In addition, records were kept of the amount of money donated within any period, although it was impossible on this measure to separate those donations made by incoming from outgoing customers.

RESULTS

The dependent variable was the percentage of donors contributing to the kettle within an observational period. That is, observational periods were assigned a percentage donor score. Shoppers within an observational period were treated as a single group, with differences between groups on percentage donor score forming the critical comparisons. The total N of the study was then the 56 observational periods, rather than the 3,703 shoppers. Since the mean group size for the Negro solicitor was 70.32 and for the Caucasian 61.93 (standard deviations equal to 53.33 and 42.98, respectively), it was assumed that the percentage score was relatively stable.

The effects of race, kettle location, and day and their interactions were analyzed by analysis of variance.

As can be seen from Table 1, both the main effect of race and of day were significant. As predicted, the Negro solicitor elicited a statistically significant lower percentage of donors than did the Caucasian. For the Negro solicitor, the average percentage donor score for observational periods was 2.22 (SD = 2.36), while for the Caucasian solicitor the average percentage donor score was 3.89 (SD = 3.60). Additionally, Saturday

shoppers were by and large less generous than Friday customers. The average percentage donor score of the group was 1.73 ($SD = 1.97$) for the Saturday shopper, and 4.38 for the Friday shopper ($SD = 3.52$).

A second dependent variable was the amount of money donated during each time period. No significant differences were found for race, day, or kettle location.

The present investigation does support, albeit equivocally, the notion that interpersonal attraction may affect donations even when the solicitors are not the eventual recipients of such contributions. While it is possible that race differences simply fail to remind observers of their social responsibilities, it is also feasible that the subjects wanted to avoid interpersonal contact with a minority group member. If this is true, then it is interesting to note that interpersonal attraction may play an important role even in those situations where personal anonymity is high and escape from unpleasant situations easy.

Discussion

The results of the first three experiments clearly replicate those of Test and Bryan and extend the findings over a variety of subject populations, settings, and tasks. The results hold for college students, motorists, and shoppers; in the university laboratory, city streets, and shopping centers; and when helping is indexed by aiding others solve arithmetic problems, changing flat tires, or donating money to the Salvation Army. The findings then are quite consistent: the presence of helping models significantly increases subsequent altruistic behavior.

That generosity breeds generosity is interesting in light of the recent concern with helping behaviors in emergency contexts. Darley and Latané[8] and Latané and Darley[9] have found that subjects are less inclined to act quickly in emergency situations when in the presence of other potential helpers. Whether faced with a medical emergency (a simulated epileptic seizure) or a dangerous natural event (simulated fire), the rapidity with which students sought to aid was reduced by the presence of others. These findings have been interpreted in three ways: as reflecting the subjects' willingness to diffuse responsibility (others will aid); as reflecting their diffusion of blame (others didn't aid either); or as reflecting conformity to the nonpanicked stooges. It is clear that the results of the first three experiments in the present series do not follow that which might be predicted by the diffusion concepts. A giving model apparently does not lend credibility to the belief that others than the self will make the necessary sacrifices. The helping other did not strengthen the observer's willingness to diffuse his social obligations, but rather stimulated greater social responsibility. In light of these results, the delayed reaction exhibited by the subjects tested by Darley and Latané might

[8] J. M. Darley, & B. Latané. Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology, 1968,* **8**, 377–383.

[9] B. Latané, & J. M. Darley. Group inhibition of bystander intervention. *Journal of Personality and Social Psychology*, 1968, **10**, 215–221.

be best attributable to conformity behavior. As they have suggested, subjects faced with a unique and stressful situation may have been either reassured by the presence of calm others or fearful of acting stupidly or cowardly. Additionally, it is possible that diffusion of responsibility is only associated with anxiety-inducing situations. The current data fail to indicate that such diffusion occurs in non-stressful situations which demand fulfillment of social obligations.

While it appears clear that the behavior of the motorists and shoppers was not dictated by a variety of situational and social pressures usually associated with the study of modeling in adults or experiments in academic settings (Orne, 1962), the mechanisms underlying the effects are not obvious. While the presence of the model in the flat-tire study may have reminded the motorists as to the social responsibility norm, a hypothesis does not appear reasonable in accounting for the results in the coins-in-the-kettle series. The bell-ringing Salvation Army worker, with kettle and self placed squarely in the pathway of the oncoming pedestrian, would seem to be reminder enough of one's obligation toward charity. A priori, it would not appear necessary to superimpose upon that scene the donating other for purposes of cognitive cueing (Wheeler, 1966).

One hypothesis to account for the model effect is that the observer is given more information regarding the consequences of such donation behavior. Experient IV suggested that solicitor status or personal attraction might operate on donation behaviors even under conditions of personal anonymity and few social constraints. It is possible that the model serves to communicate to the potential donor relevant information concerning the consequences of his act. That is, the model may demonstrate that an approach to the solicitor does not involve an unwanted interpersonal interaction (e.g., lectures on religion).

A second hypothesis to account for the data pertains to the shame-provoking capacities of the model. It is reasonable to assume that most people feel that they are, by and large, benevolent and charitable. Furthermore, it is likely that such a self-image is rarely challenged: first because charitable acts are not frequently required, second, at least in the street scenes employed in the current series of studies, solicitations are made in the context of many nongiving others. That is, a multitude of negative models—of noncharitable others—surround the solicitations in the current series of studies. Indeed, the contexts are such that most people are not helping; many more cars pass than stop to offer aid to the lady in distress; and there are many more people who refuse to put coins in the kettle than those who do. However, the witnessing of a donor, an individual who not only recognizes his social responsibility but in fact acts upon it, may produce a greater challenge to the good-self image of the observer. Acts rather than thoughts may be required of the observer in order to maintain the self-image of benevolence and charity. If such is the case, then the model characteristics most effective in producing pro-social behavior by socialized adults would be those directed toward shame

or guilt production (e.g., donations from the poor), rather than those reflecting potential reinforcement power (e.g., donations from the high status).

Whatever the mechanism underlying the model effect, it does appear quite clear that prosocial behavior can be elicited through the observation of benign others.

References

Bandura, A., Ross, D. & Ross, S. Vicarious reinforcement and imitative learning. *Journal of Abnormal and Social Psychology*, 1963, **66**, 601–607.

Bandura, A., & Walters, R. H. *Social learning and personality development*. New York: Holt, Rinehart & Winston, 1963.

Berkowitz, L. A laboratory investigation of social class and national differences in helping behavior. *International Journal of Psychology*, 1966, **1**, 231–240.

Berkowitz, L., & Daniels, L. Responsibility and dependency. *Journal of Abnormal and Social Psycholoy*, 1963, **66**, 429–436.

Berkowitz, L., & Daniels, L. Affecting the salience of the social responsibility norm: Effects of past help on the response to dependency relationship. *Journal of Abnormal and Social Psychology*, 1964, **68**, 275–281.

Berkowitz, L., & Friedman, P. Some social class differences in helping behavior. *Journal of Personality and Social Psychology*, 1967, **5**, 217–225.

Berkowitz, L., Klanderman, S. B., & Harris, R. Effects of experimenter awareness and sex of subject and experimenter on reactions to dependency relationships. *Sociometry*, 1964, **27**, 327–337.

Daniels, L., & Berkowitz, L. Liking and response to dependency relationship. *Human Relations*, 1963, **16**, 141–148.

Darlington, R. B., & Macker, C. E. Displacement of guilt-produced altruistic behavior. *Journal of Personality and Social Psychology*, 1966, **4**, 442–443.

Freed, A., Chandler, P., Mouton, J., & Blake, R. Stimulus and background factors in sign violation. *Journal of Personality*, 1955, **23**, 499.

Goranson, R., & Berkowitz, L. Reciprocity and responsibility reactions to prior help. *Journal of Personality and Social Psychology*, 1966, **3**, 227–232.

Gore, P. M., & Rotter, J. B. A personality correlate of social action. *Journal of Personality* 1963, **31**, 58–64.

Gouldner, A. The norm of reciprocity: A preliminary statement. *American Sociological Review*, 1960, **25**, 161–178.

Handlon, B. J., & Gross, P. The development of sharing behavior. *Journal of Abnormal and Social Psychology*, 1959, **59**, 425–428.

Lefkowitz, M., Blake, R., & Mouton, J. Status factors in pedestrian violation of traffic signals. *Journal of Abnormal and Social Psychology*, 1955, **51**, 704–706.

Midlarsky, E., & Bryan, J. H. Training charity in children. *Journal of Personality and Social Psychology*, 1967, **5**, 408–415.

Orne, M. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, **17**, 776–783.

Rosenbaum, M. The effect of stimulus and background factors on the volunteering response. *Journal of Abnormal and Social Psychology*, 1956, **53**, 118–121.

Rosenbaum, M., & Blake, R. Volunteering as a function of field structure. *Journal of Abnormal and Social Psychology*, 1955, **50**, 193–196.

Rosenhan, D., & White, G. M. Observation and rehearsal as determinants of prosocial behavior. *Journal of Personality and Social Psychology*, 1967, **5**, 424–431.

Ugurel-Semin, R. Moral behavior and moral

judgment of children. *Journal of Abnormal and Social Psychology,* 1952, **47**, 463–474.
Wheeler, L. Toward a theory of behavioral contagion. *Psychological Review*, 1966, **73**, 179–192.
Wright, B. A. Altruism in children and perceived conduct of others. *Journal of Abnormal and Social Psychology*, 1942, **37**, 218–233.

Exercise 9

Social Responsibility and Dependency

Why do people help each other even when the helper is unlikely to receive material or social rewards? Berkowitz and Daniels (1963) have suggested that helpers who do not directly benefit from their good deeds are conforming to a widely accepted norm of social responsibility which prescribes that individuals should help those in need. But, postulating a norm does not really explain why people help or how the norm evolved in our culture. It may be more profitable to ask, "Under what conditions are people more or less likely to help others?"

Among the many researchers who have collected data to answer this question are Latané and Darley (1970), who have consistently demonstrated that an individual is *less* likely to help when *more* potential helpers are present. For example, they have found that when subjects are alone, they are more likely to aid an epileptic (see the Research Example, p. 237) or a theft victim than when they are accompanied by at least one other potential helper. These researchers suggest that individuals diffuse responsibility in the presence of an audience so that each member of the group or crowd bears only a small share of the responsibility for helping. Perhaps an individual in a group thinks that others will help; then, if the others fail to help, he thinks "they're not helping, why should I?" This exercise is designed to test the diffusion of responsibility hypothesis which states that an individual helps less when he is in a group than when he is the only person who can help.

Procedure

Select a site where pedestrian traffic is light enough so that an unaccompanied pedestrian may witness a one-minute emergency scene without other

pedestrians entering the immediate area. Dorm hallways, residential streets, or building corridors may provide good locations. Furthermore, you must choose a site that allows you to observe the behavior of the subjects unobtrusively under "Alone" conditions.

The purpose of the experiment is to determine the impact of witnesses. In the "Crowd" condition, you will provide additional bystanders who will witness a simulated emergency. You will compare the behavior of these subject-witnesses who believe they are part of a group with other subjects who are alone with a victim.

You will need to recruit at least two collaborators to form an experimental team consisting of: (1) a recorder, (2) a victim, and (3) one or more witnesses. The recorder must locate himself out of the subject's sight in order to record the data unobtrusively. The witnesses will serve as a second (third, etc.) observer in the Crowd condition. Whenever a prospective subject approaches the scene, the victim will simulate fainting and the observer(s) will record the amount of time that elapses before help, if any, is offered.

Crowd Condition. Station the victim so that the onlookers are apparently unacquainted with the victim. As the subject approaches the victim, and reaches a predetermined distance from him, the victim comments (to no one in particular) that he feels faint, and then falls to the ground. The recorder and the witness(es) are stationed at approximately equal distances from the victim on each testing occasion. They should look at the victim when he faints without engaging in any conversation, interaction, or eye contact with either the victim or the subject.

The recorder and the witness should note the time taken for the subject either to help directly by offering aid to the victim or to help indirectly by seeking help. The experimental team must be careful that the subject does not know that he is being observed. If the subject offers to help (directly or indirectly through the solicitation of help from onlookers) his score is the number of seconds taken to *initiate* the action. When aid is offered, the victim should indicate that he (or she) is quite all right and politely decline any further assistance. If the subject simply procceds on without helping, wait a short time to see if he returns. If he does not, he is assigned an arbitrary score of 60 seconds. You may prefer a period of time shorter or longer than one minute. The witness should record his time on Data Sheet 1; the recorder on Data Sheet 2.

Alone Condition. Alone (comparison) subjects are treated identically to the Crowd (experimental) subjects except that there are no other potential helpers present to witness the fainting spell. As in the crowd condition, the observer(s) should not be visible to the potential helper.

Subjects. Since there is evidence that males and females differ in helping behavior (Schopler and Bateson, 1965), it is important to have each sex equally divided between the two conditions. Assign approximately 14 subjects to each condition and equate the number of males and females in each. You may wish, depending on your location, to run only males or only females.

Data Analysis

The data should be recorded on Data Sheets 1 and 2.

Helping. Compare the time taken to initiate rescue activity by the two groups by means of a chi-square test (Data Sheet 3 and Appendix B). Did the two groups differ in their rescue behavior and did they differ in the predicted direction?

Reliability (Optional). Did the two recorders agree in their time judgments, that is, was helping behavior reliably measured? Compute a rank-order correlation (Appendix C) between the rescue time ranks of the recorder (Column 3 of Data Sheet 2) and the witness (Column 3 of Data Sheet 1). What does it mean if the interobserver reliability is low? Is the reliability higher for the alone condition than the crowd condition? If so, why?

References

Berkowitz, L., and Daniels, L. R. "Responsibility and dependency." *Journal of Abnormal and Social Psychology*, 1963, **66**, 429–436.

Berkowitz, L., and Daniels, L. R. "Affecting the salience of the social responsibility norm: Effects of past help on the responses to dependency relationships." *Journal of Abnormal and Social Psychology*, 1964, **66**, 275–281.

These studies report the data upon which the norm of social responsibility was postulated. These experiments provided major impetus for the investigation of helping behavior in non-emergency situations.

Latané, B., and Darley, J. M. *The Unresponsive Bystander: Why Doesn't He Help?* Appleton-Century-Crofts, 1970.

Latané, B., and Darley, J. M. "Apathy." American Scientist, 1969, **57**, 244–268.

This work describes a series of studies on social determinants of rescue behavior by the two investigators most responsible for stimulating work in this area. One of these studies is included as the Research Example for this Exercise.

Schopler, J., and Bateson, N. "The power of dependence." *Journal of Personality and Social Psychology*, 1965, **2**, 247–254.

Data from this study indicate that women help a highly-dependent partner more than they help a less dependent partner. The pattern was reversed for men.

Data Sheet 1
Witness' Data Sheet

Column Numbers

(1) Subject No. and condition	(2) Sex of Subject	(3) Rescue time in seconds (witness)	(4) Rank of rescue time (witness)
1 A			
2 C			
3 C			
4 A			
5 A			
6 C			
7 C			
8 A			
9 A			
10 C			
11 C			
12 A			
13 A			
14 C			
15 C			
16 A			
17 A			
18 C			
19 C			
20 A			
21 A			
22 C			
23 C			
24 A			

Data Sheet 1
continued
on p. 232

Column Numbers

(1) Subject No. and condition	(2) Sex of Subject	(3) Rescue time in secons (witness)	(4) Rank of rescue time (witness)
25 A			
26 C			
27 C			
28 A			

Data Sheet 2
Recorder's Data Sheet

Column Numbers

(1) Subject No. and Condition A = Alone; C = Crowd.	(2) Sex of *S*.	(3) Rescue time in seconds (recorder).	(4) Rank of rescue time in seconds (recorder).	(5) Summary rescue time index from both recorder and witness. Add Column 3 on this page and Column 3 from Data Sheet #1.	(6) Rank of summary rescue time index.	(7) Enter a "+" if *S* is above median rank; Enter a "–" if below median rank.	(8) Repeat of *S*'s condition.
1 A							A
2 C							C
3 C							C
4 A							A
5 A							A
6 C							C
7 C							C
8 A							A
9 A							A
10 C							C
11 C							C
12 A							A
13 A							A
14 C							C
15 C							C
16 A							A
17 A							A
18 C							C
19 C							C
20 A							A
21 A							A
22 C							C
23 C							C
24 A							A
25 A							A
26 C							C
27 C							C
28 A							A

Median rank (the 50th percentile). 1/2 of *S*/s above this rank took shorter time; 1/2 of *S*s below this rank took longer time.

Data Sheet 3

From the information in Columns (7) and (8) of Data Sheet 2, fill in the following table:

(A)
Number of Alone *S* above time median.

*S*s in this cell will have an "A" in Column 8 of Data Sheet 2 and a "+" in Column 7.

$N_A = \boxed{}$

(B)
Number of Crowd *S*s above time median.

*S*s in this cell will have a "C" in Column 8 of Data Sheet 2 and a "+" in Column 7.

$N_B = \boxed{}$

(C)
Number of Alone *S*s below time median.

*S*s in this cell will have an "A" in Column 8 of Data Sheet 2 and a "–" in Column 7.

$N_C = \boxed{}$

(D)
Number of Crowd *S*s below time median.

*S*s in this cell will have a "C" in Column 8 of Data Sheet 2 and a "–" in Column 7.

$N_D = \boxed{}$

NOTE: The total of the four (4) numbers in the four (4) cells above should equal the total number of subjects being analyzed.

(To compute the chi-square, transfer the numbers in the boxes above to the appropriate Boxes in Appendix B).

Research Example

Bystander Intervention in Emergencies: Diffusion of Responsibility[1]

JOHN M. DARLEY AND BIBB LATANÉ

Ss overheard an epileptic seizure. They believed either that they alone heard the emergency, or that 1 or 4 unseen others were also present. As predicted the presence of other bystanders reduced the individual's feelings of personal responsibility and lowered his speed of reporting ($p < .01$). In groups of size 3, males reported no faster than females, and females reported no slower when the 1 other bystander was a male rather than a female. In general, personality and background measures were not predictive of helping. Bystander inaction in real-life emergencies is often explained by "apathy," "alienation," and "anomie." This experiment suggests that the explanation may lie more in the bystander's response to other observers than in his indifference to the victim.

Several years ago, a young woman was stabbed to death in the middle of a street in a residential section of New York City. Although such murders are not entirely routine, the incident received little public attention until several weeks later when the New York Times disclosed another side to the case: at least 38 witnesses had observed the attack—and none had even attempted to intervene. Although the attacker took more than half an hour to kill Kitty Genovese, not one of the 38 people who watched from the safety of their own apartments came out to assist her. Not one even lifted the telephone to call the police (Rosenthal, 1964).

Preachers, professors, and news commentators sought the reasons for such apparently conscienceless and inhumane lack of intervention. Their conclusions ranged from "moral decay," to "dehumanization produced by the urban environment," to "alienation," "anomie," and "existential despair." An analysis of the situation, however, suggests that factors other than apathy and indifference were involved.

A person witnessing an emergency situation, particularly such a frightening

SOURCE. *Journal of Personality and Social Psychology,* 1968, **8**, (4), 377–383.

[1] This research was supported in part by National Science Foundation Grants GS1238 and GS1239. Susan Darley contributed materially to the design of the experiment and ran the subjects, and she and Thomas Moriarty analyzed the data. Richard Nisbett, Susan Millman, Andrew Gordon, and Norma Neiman helped in preparing the tape recordings.

and dangerous one as a stabbing, is in conflict. There are obvious humanitarian norms about helping the victim, but there are also rational and irrational fears about what might happen to a person who does intervene (Milgram & Hollander, 1964). "I didn't want to get involved," is a familiar comment, and behind it lies fears of physical harm, public embarrassment, involvement with police procedures, lost work days and jobs, and other unknown dangers.

In certain circumstances, the norms favoring intervention may be weakened, leading bystanders to resolve the conflict in the direction of nonintervention. One of these circumstances may be the presence of other onlookers. For example, in the case above, each observer, by seeing lights and figures in other apartment house windows, knew that others were also watching. However, there was no way to tell how the other observers were reacting. These two facts provide several reasons why any individual may have delayed or failed to help. The responsibility for helping was diffused among the observers; there was also diffusion of any potential blame for not taking action; and finally, it was possible that somebody, unperceived, had already initiated helping action.

When only one bystander is present in an emergency, if help is to come, it must come from him. Although he may choose to ignore it (out of concern for his personal safety, or desires "not to get involved"), any pressure to intervene focuses uniquely on him. When there are several observers present, however, the pressures to intervene do not focus on any one of the observers; instead the responsibility for intervention is shared among all the onlookers and is not unique to any one. As a result, no one helps.

A second possibility is that potential blame may be diffused. However much we may wish to think that an individual's moral behavior is divorced from considerations of personal punishment or reward, there is both theory and evidence to the contrary (Aronfreed, 1964; Miller & Dollard, 1941; Whiting & Child, 1953). It is perfectly reasonable to assume that, under circumstances of group responsibility for a punishable act, the punishment or blame that accrues to any one individual is often slight or nonexistent.

Finally, if others are known to be present, but their behavior cannot be closely observed, any one bystander can assume that one of the other observers is already taking action to end the emergency. Therefore, his own intervention would be only redundant—perhaps harmfully or confusingly so. Thus, given the presence of other onlookers whose behavior cannot be observed, any given bystander can rationalize his own inaction by convincing himself that "somebody else must be doing something."

These considerations lead to the hypothesis that the more bystanders to an emergency, the less likely, or the more slowly, any one bystander will intervene to provide aid. To test this proposition it would be necessary to create a situation in which a realistic "emergency" could plausibly occur. Each subject should also be blocked from communicating with others to prevent his getting information about their

behavior during the emergency. Finally, the experimental situation should allow for the assessment of the speed and frequency of the subjects' reaction to the emergency. The experiment reported below attempted to fulfill these conditions.

Procedure

Overview

A college student arrived in the laboratory and was ushered into an individual room from which a communication system would enable him to talk to the other participants. It was explained to him that he was to take part in a discussion about personal problems associated with college life and that the discussion would be held over the intercom system, rather than face-to-face, in order to avoid embarrassment by preserving the anonymity of the subjects. During the course of the discussion, one of the other subjects underwent what appeared to be a very serious nervous seizure similar to epilepsy. During the fit it was impossible for the subject to talk to the other discussants or to find out what, if anything, they were doing about the emergency. The dependent variable was the speed with which the subjects reported the emergency to the experimenter. The major independent variable was the number of people the subject thought to be in the discussion group.

Subjects

Fifty-nine female and thirteen male students in introductory psychology courses at New York University were contacted to take part in an unspecified experiment as part of a class requirement.

Method

Upon arriving for the experiment, the subject found himself in a long corridor with doors opening off it to several small rooms. An experimental assistant met him, took him to one of the rooms, and seated him at a table. After filling out a background information form, the subject was given a pair of headphones with an attached microphone and was told to listen for instructions.

Over the intercom, the experimenter explained that he was interested in learning about the kinds of personal problems faced by normal college students in a high pressure, urban environment. He said that to avoid possible embarrassment about discussing personal problems with strangers several precautions had been taken. First, subjects would remain anonymous, which was why they had been placed in individual rooms rather than face-to-face. (The actual reason for this was to allow tape recorder simulation of the other subjects and the emergency.) Second, since the discussion might be inhibited by the presence of outside listeners, the experimenter would not listen to the initial discussion, but would get the subject's reactions later, by questionnaire. (The real purpose of this was to remove the obviously responsible experimenter from the scene of the emergency.)

The subjects were told that since the experimenter was not present, it was

necessary to impose some organization. Each person would talk in turn, presenting his problems to the group. Next, each person in turn would comment on what the others had said, and finally, there would be a free discussion. A mechanical switching device would regulate this discussion sequence and each subject's microphone would be on for about 2 minutes. While any microphone was on, all other microphones would be off. Only one subject, therefore, could be heard over the network at any given time. The subjects were thus led to realize when they later heard the seizure that only the victim's microphone was on and that there was no way of determining what any of the other witnesses were doing, nor of discussing the event and its possible solution with the others. When these instructions had been given, the discussion began.

In the discussion, the future victim spoke first, saying that he found it difficult to get adjusted to New York City and to his studies. Very hesitantly, and with obvious embarrassment, he mentioned that he was prone to seizures, particularly when studying hard or taking exams. The other people, including the real subjects, took their turns and discussed similar problems (minus, of course, the proneness to seizures). The naive subject talked last in the series, after the last prerecorded voice was played.[2]

When it was again the victim's turn to talk, he made a few relatively calm comments, and then, growing increasingly louder and incoherent, he continued:

> I-er-um-I think I-I need-er-if-if could-er-er-somebody er-er-er-er-er-er-er give me a little-er-give me a little help here because-er-I-er-I'm-er-er-h-h-having a-a-a real problem-er-right now and I-er-if somebody could help me out it would-it would-er-er- s-s-sure be-sure be good . . . because-er-there-er-er-a cause I-er-I-uh-I've got a-a one of the-er-sei-----er-er-things coming on and-and-and I could really-er-use some help so if somebody would-er-give me a little h-help-uh-er-er-er-er-er c-could somebody-er-er-er-er-help--er-uh-uh-uh (choking sounds). . . . I'm gonna die--er-help-er-er-seizure-er [chokes, then quiet].

The experimenter began timing the speed of the real subject's response at the beginning of the victim's speech. Informed judges listening to the tape have estimated that the victim's increasingly louder and more disconnected ramblings clearly represented a breakdown about 70 seconds after the signal for the victim's second speech. The victim's speech was abruptly cut off 125 seconds after this signal, which could be interpreted by the subject as indicating that the time allotted for that speaker had elapsed and the switching circuits had switched away from him. Times reported in the results are measured from the start of the fit.

Group Size Variable

The major independent variable of the study was the number of other people

[2] To test whether the order in which the subjects spoke in the first discussion round significantly affected the subjects' speed of report, the order in which the subjects spoke was varied (in the six-person group). This had no significant or noticeable effect on the speed of the subjects' reports.

that the subject believed also heard the fit. By the assistant's comments before the experiment, and also by the number of voices heard to speak in the first round of the group discussion, the subject was led to believe that the discussion group was one of three sizes: either a two-person group (consisting of a person who would later have a fit and the real subject), a three-person group (consisting of the victim, the real subject, and one confederate voice), or a six-person group (consisting of the victim, the real subject, and four confederate voices). All the confederates' voices were tape-recorded.

Variations in Group Composition

Varying the kind as well as the number of bystanders present at an emergency should also vary the amount of responsibility felt by any single bystander. To test this, several variations of the three-person group were run. In one three-person condition, the taped by-stander voice was that of a female, in another a male, and in the third a male who said that he was a premedical student who occasionally worked in the emergency wards at Bellevue hospital.

In the above conditions the subjects were female college students. In a final condition males drawn from the same introductory psychology subject pool were tested in a three-person female-bystander condition.

Time to Help

The major dependent variable was the time elapsed from the start of the victim's fit until the subject left her experimental cubicle. When the subject left her room, she saw the experimental assistant seated at the end of the hall, and invariably went to the assistant. If 6 minutes elapsed without the subject having emerged from her room, the experiment was terminated.

As soon as the subject reported the emergency, or after 6 minutes had elapsed, the experimental assistant disclosed the true nature of the experiment, and dealt with any emotions aroused in the subject. Finally the subject filled out a questionnaire concerning her thoughts and feelings during the emergency, and completed scales of Machiavellianism, anomie, and authoritarianism (Christie, 1964), a social desirability scale (Crowne & Marlowe, 1964), a social responsibility scale (Daniels & Berkowitz, 1964), and reported vital statistics and socioeconomic data.

Results

PLAUSIBILITY OF MANIPULATION

Judging by the subjects' nervousness when they reported the fit to the experimenter, by their surprise when they discovered that the fit was simulated, and by comments they made during the fit (when they thought their microphones were off), one can conclude that almost all of the subjects perceived the fit as real. There were two exceptions in different experimental conditions, and the data for these subjects were dropped from the analysis.

EFFECT OF GROUP SIZE ON HELPING

The number of bystanders that the subject perceived to be present had a

Table 1

Effects of Groups Size on Likelihood and Speed of Response

Group size	N	% responding by end of fit	Time in sec.	Speed score
2 (S & victim)	13	85	52	.87
3 (S, victim, & 1 other)	26	62	93	.72
6 (S, victim, & 4 others)	13	31	166	.51

Note.—p value of differences: $x^2 = 7.91, p < .02$,[a] $F = 8.09, p < .01$, for speed scores.
[a]Notations such as $p < .02$ are explained in the Introduction, pp. 15–16.

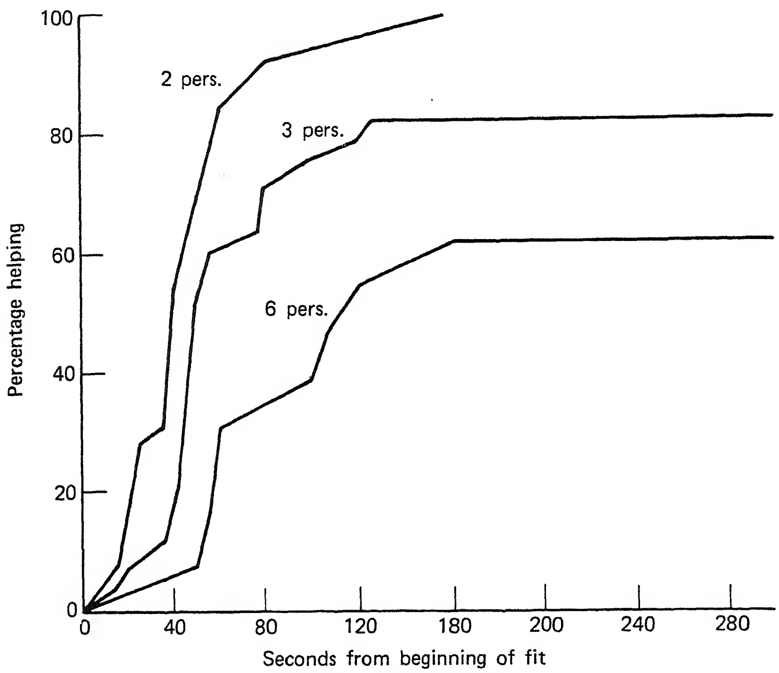


Fig. 1. Cumulative distributions of helping responses.

major effect on the likelihood with which she would report the emergency (Table 1). Eighty-five percent of the subjects who thought they alone knew of the victim's plight reported the seizure before the victim was cut off, only 31% of those who thought four other bystanders were present did so.

Every one of the subjects in the two-person groups, but only 62% of the subjects in the six-person groups, ever reported the emergency. The cumulative distributions of response times for groups of different perceived size (Figure 1) indicates that, by any point in time, more subjects from the two-person

Table 2

Effects of Group Composition on Likelihood and Speed of Response[a]

Group composition	N	% responding by end of fit	Time in sec.	Speed score
Female S, male other	13	62	94	74
Female S, female other	13	62	92	71
Female S, male medic other	5	100	60	77
Male S, female other	13	69	110	68

[a]Three-person group, male victim.

groups had responded than from the three-person groups, and more from the three-person groups than from the six-person groups.

Ninety-five percent of all the subjects who ever responded did so within the first half of the time available to them. No subject who had not reported within 3 minutes after the fit ever did so. The shape of these distributions suggest that had the experiment been allowed to run for a considerably longer time, few additional subjects would have responded.

SPEED OF RESPONSE

To achieve a more detailed analysis of the results, each subject's time score was transformed into a "speed" score by taking the reciprocal of the response time in seconds and multiplying by 100. The effect of this transformation was to deemphasize differences between longer time scores, thus reducing the contribution to the results of the arbitrary 6-minute limit on scores. A high speed score indicates a fast response.

An analysis of variance indicates that the effect of group size is highly significant ($p < .01$). Duncan multiple-range tests indicate that all but the two- and three-person groups differ significantly from one another ($p < .05$).

VICTIM'S LIKELIHOOD OF BEING HELPED

An individual subject is less likely to respond if he thinks that others are present. But what of the victim? Is the inhibition of the response of each individual strong enough to counteract the fact that with five onlookers there are five times as many people available to help? From the data of this experiment, it is possible mathematically to create hypothetical groups with one, two, or five observers.[3] The calculations indicate that the victim is about equally likely to get help from one bystander as from two. The victim is considerably more likely to have gotten help from one or two observers than from five during the first minute of the fit. For instance, by 45 seconds after the start of the fit, the

[3]The formula for the probability that at least one person will help by a given time is $1-(1-P)^n$ where n is the number of observers and P is the probability of a single individual (who thinks he is one of n observers) helping by that time.

victim's chances of having been helped by the single bystanders were about 50%, compared to none in the five observer condition. After the first minute, the likelihood of getting help from at least one person is high in all three conditions.

EFFECTS OF GROUP COMPOSITION ON HELPING THE VICTIM

Several variations of the three-person group were run. In one pair of variations, the female subject thought the other bystander was either male or female; in another, she thought the other bystander was a premedical student who worked in an emergency ward at Bellevue hospital. As Table 2 shows, the variations in sex and medical competence of the other bystander had no important or detectable effect on speed of response. Subjects responded equally frequently and fast whether the other bystander was female, male, or medically experienced.

SEX OF THE SUBJECT AND SPEED OF THE RESPONSE

Coping with emergencies is often thought to be the duty of males, especially when females are present, but there was no evidence that this was the case in this study. Male subjects responded to the emergency with almost exactly the same speed as did females (Table 2).

REASONS FOR INTERVENTION OR NONINTERVENTION

After the debriefing at the end of the experiment each subject was given a 15-item checklist and asked to check those thoughts which had "crossed your mind when you heard Subject 1 calling for help." Whatever the condition, each subject checked very few thoughts, and there were no significant differences in number or kind of thoughts in the different experimental groups. The only thoughts checked by more than a few subjects were "I didn't know what to do" (18 out of 65 subjects), "I thought it must be some sort of fake" (20 out of 65), and "I didn't know exactly what was happening" (26 out of 65).

It is possible that subjects were ashamed to report socially undesirable rationalizations, or, since the subjects checked the list *after* the true nature of the experiment had been explained to them, their memories might have been blurred. It is our impression, however, that most subjects checked few reasons because they had few coherent thoughts during the fit.

We asked all subjects whether the presence or absence of other bystanders had entered their minds during the time that they were hearing the fit. Subjects in the three- and six-person groups reported that they were aware that other people were present, but they felt that this made no difference to their own behavior.

INDIVIDUAL DIFFERENCE CORRELATES OF SPEED OF REPORT

The correlations between speed of report and various individual differences on the personality and background measures were obtained by normalizing the distribution of report speeds within each experimental condition and pooling these scores across all conditions (n =

62–65). Personality measures showed no important or significant correlations with speed of reporting the emergency. In fact, only one of the 16 individual difference measures, the size of the community in which the subject grew up, correlated ($r = -.26, p < .05$) with the speed of helping.

Discussion

Subjects, whether or not they intervened, believed the fit to be genuine and serious. "My God, he's having a fit," many subjects said to themselves (and were overheard via their microphones) at the onset of the fit. Others gasped or simply said "Oh." Several of the male subjects swore. One subject said to herself, "It's just my kind of luck, something has to happen to me!" Several subjects spoke aloud of their confusion about what course of action to take, "Oh God, what should I do?"

When those subjects who intervened stepped out of their rooms, they found the experimental assistant down the hall. With some uncertainty, but without panic, they reported the situation. "Hey, I think Number 1 is very sick. He's having a fit or something." After ostensibly checking on the situation, the experimenter returned to report that "everything is under control". The subjects accepted these assurances with obvious relief.

Subjects who failed to report the emergency showed few signs of the apathy and indifference thought to characterize "unresponsive bystanders." When the experimenter entered her room to terminate the situation, the subject often asked if the victim was "all right." "Is he being taken care of?" "He's all right isn't he?" Many of these subjects showed physical signs of nervousness; they often had trembling hands and sweating palms. If anything, they seemed more emotionally aroused than did the subjects who reported the emergency.

Why, then, didn't they respond? It is our impression that nonintervening subjects had not decided *not* to respond. Rather they were still in a state of indecision and conflict concerning whether to respond or not. The emotional behavior of these nonresponding subjects was a sign of their continuing conflict, a conflict that other subjects resolved by responding.

The fit created a conflict situation of the avoidance-avoidance type. On the one hand, subjects worried about the guilt and shame they would feel if they did not help the person in distress. On the other hand, they were concerned not to make fools of themselves by overreacting, not to ruin the ongoing experiment by leaving their intercom, and not to destroy the anonymous nature of the situation which the experiment had earlier stressed as important. For subjects in the two-person condition, the obvious distress of the victim and his need for help were so important that their conflict was easily resolved. For the subjects who knew there were other bystanders present, the cost of not helping was reduced and the conflict they were in more acute. Caught between the two negative alternatives of letting the victim continue to suffer or the costs of rushing in to help, the nonresponding bystanders

vacillated between them rather than choosing not to respond. This distinction may be academic for the victim, since he got no help in either case, but it is an extremely important one for arriving at an understanding of the causes of bystanders' failures to help.

Although the subjects experienced stress and conflict during the experiment, their general reactions to it were highly positive. On a questionnaire administered after the experimenter had discussed the nature and purpose of the experiment, every single subject found the experiment either "interesting" or "very interesting" and was willing to participate in similar experiments in the future. All subjects felt they understood what the experiment was about and indicated that they thought the deceptions were necessary and justified. All but one felt they were better informed about the nature of psychological research in general.

Male subjects reported the emergency no faster than did females. These results (or lack of them) seem to conflict with the Berkowitz, Klanderman, and Harris (1964) finding that males tend to assume more responsibility and take more initiative than females in giving help to dependent others. Also, females reacted equally fast when the other bystander was another female, a male, or even a person practiced in dealing with medical emergencies. The ineffectiveness of these manipulations of group composition cannot be explained by general insensitivity of the speed measure, since the group-size variable had a marked effect on report speed.

It might be helpful in understanding this lack of difference to distinguish two general classes of intervention in emergency situations: direct and reportorial. Direct intervention (breaking up a fight, extinguishing a fire, swimming out to save a drowner) often requires skill, knowledge, or physical power. It may involve danger. Anerican cultural norms and Berkowitz's results seem to suggest that males are more responsible than females for this kind of direct intervention.

A second way of dealing with an emergency is to report it to someone qualified to handle it, such as the police. For this kind of intervention, there seem to be no norms requiring male action. In the present study, subjects clearly intended to report the emergency rather than take direct action. For such indirect intervention, sex or medical competence does not appear to affect one's qualifications or responsibilities. Anybody, male or female, medically trained or not, can find the experimenter.

In this study, no subject was able to tell how the other subjects reacted to the fit. (Indeed, there were no other subjects actually present.) The effects of group size on speed of helping, therefore, are due simply to the perceived presence of others rather than to the influence of their actions. This means that the experimental situation is unlike emergencies, such as a fire, in which bystanders interact with each other. It is, however, similar to emergencies, such as the Genovese murder, in which spectators knew others were also watching but were prevented by walls between them from communication that might have counteracted the diffusion of responsibility.

The present results create serious difficulties for one class of commonly given

explanations for the failure of bystanders to intervene in actual emergencies, those involving apathy or indifference. These explanations generally assert that people who fail to intervene are somehow different in kind from the rest of us, that they are "alienated by industrialization," "dehumanized by urbanization," "depersonalized by living in the cold society," or "psychopaths." These explanations serve a dual function for people who adopt them. First, they explain (if only in a nominal way) the puzzling and frightening problem of why people watch others die. Second, they give individuals reason to deny that they too might fail to help in a similar situation.

The results of this experiment seem to indicate that such personality variables may not be as important as these explanations suggest. Alienation, Machiavellianism, acceptance of social responsibility, need for approval, and authoritarianism are often cited in these explanations. Yet they did not predict the speed or likelihood of help. In sharp contrast, the perceived number of bystanders did. The explanation of bystander "apathy" may lie more in the bystander's response to other observers than in presumed personality deficiencies of "apathetic" individuals. Although this realization may force us to face the guilt-provoking possibility that we too might fail to intervene, it also suggests that individuals are not, of necessity, "noninterveners" because of their personalities. If people understand the situational forces that can make them hesitate to intervene, they may better overcome them.

References

Aronfreed, J. The origin of self-criticism. *Psychological Review,* 1964, **71**, 193-219.

Berkowitz, L., Klanderman, S., & Harris, R. Effects of experimenter awareness and sex of subject on reactions to dependency relationship. *Sociometry,* 1964, **27**, 327–329.

Christie, R. The prevalence of machiavellian orientations. Paper presented at the meeting of the American Psychological Association, Los Angeles, 1964.

Crowne, D., & Marlowe, D. *The approval motive.* New York: Wiley, 1964.

Daniels, L., & Berkowitz, L. Liking and response to dependency relationships. *Human Relations* 1963, **16**, 141–148.

Milgram, S., & Hollander, P. Murder they heard. *Nation* 1964, **198**, 602–604.

Miller, N., & Dollard, J. *Social learning and imitation.* New Haven: Yale University Press, 1941.

Rosenthal, A.M. *Thirty-eight witnesses.* New York: McGraw-Hill, 1964.

Whiting, J. W. M., & Child, I. *Child training and personality.* New Haven: Yale University Press, 1953.

Exercise 10

Selective Exposure to Information

How do you decide whether to watch TV or to read the newspaper or to browse through a magazine? What magazines do you choose to read and how closely do you read them? Which articles attract your attention in the newspaper? Obviously it is impossible for people in a modern society to process all of the information available to them. How, then, does a person select what to read and listen to? The selection process is certainly not a random one. We know that we quite consciously choose to subscribe to certain magazines and newspapers and to attend certain public lectures. But sometimes choices about what kinds of information to attend to are less conscious. Why is our attention drawn to one newspaper article and not another?

Several authors have suggested that we seek certain kinds of information and avoid others. For example, Festinger in his *A Theory of Cognitive Dissonance* (1957, pp. 3, 123 ff, 162ff) has suggested that we selectively expose ourselves to supportive information that is likely to be consistent with our behaviors and beliefs, and selectively avoid information that is dissonant or psychologically inconsistent. According to this theory, a smoker would be likely to avoid an article connecting lung cancer to smoking because the information that smoking causes cancer could be disturbing when linked with the knowledge that he is a smoker. On the other hand, a person who has just managed to stop smoking would be especially likely to seek out and read the same article.

In addition to problems of differential attention, there may be memory differences for various kinds of material. Why do we remember one TV commercial and not others? If you were asked to recall what you read

several days later, which parts of an article might you remember? Would your memory be more accurate for some kinds of information than for others?

Numerous field and laboratory experiments have been conducted to test dissonance-theory ideas about selective exposure and remembering (see Freedman and Sears, 1965, for a review), and although some investigators have found supporting evidence, it is now clear that the determinants of attending, selecting, and remembering are quite complex. Under some circumstances, people even prefer to read dissonant information which is useful, novel, or relevant.

For example, Canon (1964) has found that subjects will choose to read painfully dissonant material when it is likely to be useful. A car owner may not want to read about defects in his new automobile, but he may seek out such information if he can use it to correct safety hazards. Or consider a devout precinct worker for candidate X who attends candidate Y's rally. The information he hears may be useful in the campaign.

Receptivity to dissonant information may also be influenced by how much supportive information has already been assimilated (Brock and Balloun, 1967). A bored person can be attracted to novel or exciting information even if it is inconsistent with his beliefs. The precinct worker may find opposition arguments unconvincing, but interesting because they are unfamiliar.

In addition to differences in the utility or novelty of information which qualify the dissonance selective exposure hypothesis, it has been suggested that there are wide individual differences in preference for information. For example, confident, high self-esteem, or politically active people may be more willing to undergo the risks and discomfort associated with exposure to nonsupportive or threatening information.

In this exercise you will have an opportunity to study how information presented in newspaper articles may be processed, assimilated, and remembered according to reader motives; emphasis is especially on the motive to maintain cognitive consistency by seeking supportive information or avoiding dissonant information. Assimilation of information can be affected at three points: (1) selective attention—noticing some articles and not others, (2) selective exposure—reading only some of the noticed articles, and (3) selective forgetting or distortion of facts presented in the article. Thus, a particular subject may not notice an article, notice it but not read it, or read it but forget it. Brock and Balloun (1967), for example, have attempted to distinguish between just noticing information and more cognitively deciding whether to read (listen to) it carefully.

This exercise concerns peoples' reactions to a newspaper article about smoking and lung cancer.[1] You will attempt to determine the extent to which attending, avoiding, forgetting, and distorting of this article are a function of the motives of the smoking and nonsmoking readers.

Instructions

Appended at the end of this exercise on pages 257-263 are 11 articles that were copied from newspapers. Remove these pages, clip out the articles, and place them in order (number 1 on the left, number 11 on the right) on an empty table in a room of minimal distractions. The critical article on smoking should be ninth from the left. Recruit a subject to "help you standardize some very brief tests you are working on." Later you will explain the purpose of the experiment, but if you tell him in advance it may not be possible for you to obtain his natural reactions. Have the subject put down any books he may be carrying and lead him to the table where you have placed the articles. As you move the articles aside make a casual remark such as "These are reprinted newspaper articles from a journalism class." Present the subject with the "College Attitude Survey," which you will find on page 265. Tell him that this is the first of two very brief attitude tests and that you will return with the second one in ten minutes. Then leave the room and return in exactly ten minutes. The Survey can be completed easily in three or four minutes; thus the procedure will usually allow enough excess time for the subject to browse through the clippings. Make sure that the subject cannot return to his own books or other materials during this time. When you return, collect the questionnaire, remove the articles from the subject's vision and inform him that you will not give him a second questionnaire because the true purpose of the experiment is to determine what kinds of information people notice, read, and remember. Ask him if he has looked through the articles on the table if it is not obvious. If you find that the first several subjects you test did not read the articles, you can change the procedures by lengthening the time you are out of the room, or by introducing an off-hand comment upon leaving such as "I might be a few minutes getting the other questionnaire—you can look at those articles if you finish the questionnaire." If you find that the first several subjects read all of the articles, you will need to shorten the time that they are left alone. For

[1] The article which asserts that women smokers are *less* likely to contract lung cancer than nonsmokers was published February 23, 1966 in the *San Francisco Chronicle*, page 1; however, to the disappointment of women smokers, it was an error and a correction appeared the next day.

subjects who have not looked at the articles, ascertain smoking habits by administering the smoking questionnaire on page 273, and terminate the interview. For subjects who have read at least some of the articles, first administer the recall test on page 267, and then the recognition test on page 269. Two items, numbers 6 and 12, were not included in the stimulus material, but appear on the recognition test as a check against subjects who present themselves falsely, are careless, or are susceptible to responding in a certain way without regard to the item (response bias). Subjects who claim that they recognize either of these articles should be excluded from the data analysis. Next, ask those subjects who either recall or recognize the critical story on smoking to answer the comprehension questionnaire on page 271 and, finally, all subjects should fill in the smoking questions on page 273. You may want to tear out the convenient procedure summary on page 283 and refer to it during the experiment.

Unless your subject smokes a cigarette in your presence, you will usually not know how to classify him until after you administer the smoking questionnaire. This procedural order is intentional to avoid sensitizing the subject to the topic of interest, and to prevent experimenter bias when interacting with the subject. The final questionnaire divides subjects into two groups of smokers (content or want to quit) and two groups of nonsmokers (quitters or never-smoked). Continue to run subjects until you have at least three or four in each category. Motivational theories like dissonance predict that content smokers and quitter nonsmokers will be more likely to select and bias smoking information (in opposite directions) than those who have never smoked and unhappy smokers.

Be sure to answer any questions that the subject asks, and inform him that the newspaper articles are not current (they were clipped from the February 23, 1966 *San Francisco Chronicle*).

Data Analysis

Record the recall and recognition data on the Data Recording Sheet, page 275. Since recognition is invariably greater than recall, we may assume that the articles recalled were more closely attended than those recognized but not recalled. Are there differences between groups in exposure as measured by percentage recall and percentage recognition scores? Are there differences only in attention indicated by headline or lead reading? Differences between subgroups of smokers and nonsmokers can be tested by means of chi-square tests (Appendix B). For example, the table on page 253 compares quitters and never smoked on exposure to the critical article.

Similar comparisons can be made for all smokers versus all nonsmokers. It is possible that any differences between categories of smokers and nonsmokers are because of general perceptual ability, for example, smokers may

	Quitters	Never Smoked
Recalled or Recognized #9	20	12
Did *not* Recall or Recognize #9	4	11

$\chi^2 = 3.91$ $p < .05$

read more or less articles on any subject. To rule out this possibility, record data for several articles selected as controls. Data forms are provided on pages 277-279. If the data pattern for the control articles are similar to trends for the critical article, a selective exposure hypothesis cannot be accepted. An excellent control procedure would be to combine responses to all of the noncritical articles to obtain percentage recall and percentage recognition scores for comparison with the critical article.

Comprehension responses are to be recorded on page 281. The percentage of distorters and correct perceivers can be calculated and compared by chi-square analysis.

For example:

	Smokers	Nonsmokers
Responded 2a (distortion)	11	3
Responded 2b (correct)	5	15
Responded 2c, d, e (incorrect)	8	1

$\chi^2 = 14.63$ $p < .001$[2]

It would be expected that quitters would be least likely and smokers most likely to distort the unexpected finding that women smokers are LESS likely to contract cancer than women who do not smoke. How might the pattern of results you obtained differ if the critical article for this exercise had contained the expected positive relationship between cancer and smoking?

[2]Notations such as $p < .001$ and $p < .05$ are explained in the Introduction, p. 15. The instructions for calculating chi-square (X^2) that appear in Appendix B are appropriate only for 2 X 2 tables such as the table that compares Quitters and Never-Smoked on recall. To calculate chi-square for larger tables such as the 2 X 3 or 6-cell table comparing Smokers and Nonsmokers, consult your instructor or a basic statistics text.

References

Brock, T. C., and Balloun, J. L. Behavioral receptivity to dissonant information. *Journal of Personality and Social Psychology,* 1967, **6**, 413–428.

This article reports four experiments designed to study effects of novelty, utility, and relevance on the desire to hear various messages. For example, more smokers than nonsmokers pressed a button to remove static from a message disputing the link between smoking and lung cancer. The authors conclude that dissonance theory adequately explains selective *attention*, but that other factors must be taken into account to predict selective *exposure* to future nonsupportive messages.

Canon, L. K. Self-confidence in selective exposure to information. In Festinger, L. (Ed.), *Conflict, decision and dissonance*. Stanford, California: Stanford University Press, 1964, pp. 83–95.

Canon's subjects made a decision about a business case and then chose articles either supportive or nonsupportive of their decision. Subjects preferred useful information whether or not it was supportive, and half of the subjects who had been told that their previous decisions were correct (high-confidence condition) preferred dissonant information. This latter result was not replicated in a subsequent study (see Freedman and Sears below).

Festinger, L. *A theory of cognitive dissonance*. Stanford: Stanford University Press, 1957, pp. 123–176.

Chapters 6 and 7 in this classic work present theory and data relating to voluntary and involuntary exposure to information.

Freedman, J. L., and Sears, D. O. Selective exposure. In L. Berkowitz (Ed.) *Advances in experimental social psychology,Vol. 2*. New York: Academic Press, 1965, pp. 57–97.

This chapter is the most extensive review of the selective exposure literature currently available. The Research Example is compared with other studies on p. 62 of this review.

Rhine, R. J. The 1964 presidential election and curves of information seeking and avoiding. *Journal of Personality and Social Psychology*, 1967, **5**, 416-423.

Festinger has suggested that seeking of consonant and avoiding of dissonant information are greatest when there is a moderate magnitude of dissonance; that is, people

with almost no dissonance and those with great dissonance will be relatively indifferent to new information. Rhine's study, which deals with information-seeking and avoiding during the 1964 Johnson-Goldwater Presidential campaign, partially verifies this curvilinear relationship.

Clip the following reprinted newspaper articles and display them as described on p. 251. Do not clip the numbers; they are for your reference only.

(1)

MARIJUANA'S HARM PROOF IS LACKING

Washington (AP)—Dr. Philip Handler, chairman of the National Science Board says there is no scientific proof that smoking marijuana is addictive or leads to the taking of harder drugs.

"The evidence that marijuana is addictive is not conclusive at all—quite the opposite as far as I can make out," Handler told a House appropriations subcommittee in testimony released Friday.

"It is our Puritan ethics which say we shouldn't do this rather than science which says we should not, at the moment," Handler said.

Handler, a professor and chairman of the department of biochemistry at Duke University, heads the board that sets the policies of the National Science Foundation. He is also the newly named president of the National Academy of Sciences.

Handler said it remains a scientific problem to decide whether marijuana is dangerous to society.

"Very obviously," he said, "if alcohol were discovered tomorrow for the first time we would consider it exactly in the same light and wonder whether or not our society should accept it."

At the moment, he said, the scientific evidence is inadequate to make any firm decisions about the physical or mental long-term damage or addiction resulting from smoking marijuana.

He said there is also no scientific evidence that using marijuana will lead a person to harder drugs.

(2)

NATION'S FIRST MARRIED DEACON HARD-NOSED ABOUT MILITARY LIFE

by Kim Larsen

(NC News Service)

St. Paul—America's first married deacon, himself a former Royal air force chaplain, blasted rebellion against military service and expressed doubt about the motive of anti-war protesters.

Deacon Cole said the Royal air force "made me a man and helped me to become a Christian."

A native of England, Deacon Cole served in his country's air force as a military policeman in Baghdad where, he said, he was converted "in the Billy Graham sense of the word. I had a meeting with Christ."

To be a real Christian, he told his audience, he felt he had to be active. So he became an Episcopalian priest. After three years in parish work in England, he went back to the Royal air force as a chaplain.

"I cannot understand the rebellion in this country (America) against serving in the armed forces nor the burning of draft cards," Deacon Cole said.

He said he agrees with a slogan he recently saw which said, "If you don't like America, get out." He said the slogan, "My country, right or wrong," can be carried to extremes, but admitted he sees some truth to it.

In doubting the motives of American young men in refusing to be inducted into military service, Deacon Cole said he thinks the real motive is "they do not want to get haircuts." The deacon sports a military-type haircut.

(3)

LAWYERS REJECT LEGALIZED DRUGS

Menominee, Mich. —AP— A proposal calling for legalized marijuana for adults has been flatly rejected by the board of commissioners of the State Bar of Michigan.

Meeting in conjunction with the annual Upper Peninsula Legal institute last week, the board unanimously voted down the resolution recommended by the bar's civil liberties commission, which had based its resolution on a similar stand by the American Civil Liberties Union.

The ACLU reasoned the ban on marijuana and the associated stiff criminal penalties violated individual constitutional rights.

Michigan law makes possession of the drug a felony.

(4)

FUNERAL USE PROTESTED 'MACABRE' JOBS FOR ORPHANS

Turin, Italy

A judge made a public protest yesterday against the Italian practice of hiring children from orphanages to pray at funerals.

Emilio Germano, president of the Turin civil tribunal, said it should be forbidden.

He said orphanage children were already deprived of ordinary family life and should not be inflicted with the nightmare of continually attending "sad, even macabre, anguishing scenes."

Writing in the newspaper La Stampa, the judge denounced as unjust and inhuman the spectacle of a double file of orphans, "long or short in relation to the wealth of the departed one, sent to follow funeral processions."

He blamed the state for not granting enough money to run orphanages decently.

He also chastised the "insensibility" of orphanage directors and the "exhibitionist mania" of bereaved families who were not content with a simple funeral but wanted "processions of poor babies forced to recite dreary dirges which they do not understand."

Reuters

(5)

FRENCH PLAN FOR FALLOUT

Papeete, Tahiti

French Defense Minister Pierre Messmer said yesterday that steps will be taken to assure that radioactive fallout from a French hydrogen bomb test will harm no one.

Messmer, speaking at a news conference during an inspection tour, gave no date for the firing, which is expected later this year.

Associated Press

(6)

ECUADOR FREES FOUR U.S. BOATS

Quito, Ecuador (AP)—The Ecuadorean navy seized four United States tuna boats Friday in the Gulf of Guayaquil but released them at sea, naval authorities announced.

Word of the release was welcomed in Washington, where there had been reports that more than a half-dozen boats had been seized. Some U.S. congressmen have been calling for retaliation against seizure of American vessels off South America's Pacific Coast.

Ecuadorean naval headquarters in Quito named the boats as the Royal Pacific, Neptune, Bold Venture, and Marietta, but refused to say why they were released. U.S. diplomats were said to have made stiff complaints to Ecuadorean government officials over the incident.

The Ecuadorean navy said a patrol craft intercepted the four vessels about 110 miles southwest of Guayaquil, Ecuador's main seaport. It added that there were at least 13 American boats in the area.

Reports received in Washington said the four boats had been released without damage and no fines had been imposed. In recent seizures, captains of U.S. fishing vessels have been forced to pay stiff fines.

The fishing boat seizures stem from a dispute over territorial water limits off the South American coast. Peru, Ecuador, and Chile claim 200-mile limits and the U.S. recognizes only 12-mile limits.

(7)

BEST COIN PROTECTION

Problem of Guard Against Tarnish Is Big Among Collectors

By NORMAN M. DAVIS, Journal Special Correspondence

One of the coin collector's goals is protection of his coins against sulphur, oxygen and other damaging substances. The ideal container would be airtight and chemically inert, but I dislike airtight containers that seal coins off from collectors.

I asked Dr. and Mrs. V. Clain-Stefanelli what they use at the Smithsonian Institution. The numismatics division curator showed me sulphur-free open, black cardboard trays mounted on a cardboard base. But these take more space than most collectors can afford.

Vinyl envelopes? One dealer hears "comments both for and against." Mrs. Stefanelli showed me a greasy, smelly deposit that formed on vinyl holders within months.

Pages with transparent slides showing both sides of coins form "the most popular type" coin album, reports a dealer.

Coin mounts and albums, both cardboard, can tarnish coin edges after some years.

For mounts, a dealer suggests using larger ones than necessary, to provide space around coins. Some dealers recommend using albums only for circulated coins.

Clear plastic holders are popular with many collectors.

What about 2 by 2 paper coin envelopes? "Not for uncirculated coins!" says one dealer. "Exceptionally high sulphur," charges another. But some dealers call coin envelopes relatively low in sulphur. The Stefanellis sometimes use such envelopes temporarily, without visible damage.

Two dealers suggest what may be the simplest, safest and most economical system:

First, put coins in inert envelopes.

Then put the coins in standard coin envelopes. Your coins are protected and you still have a place to write information.

(8)

SHRIMP CAPTAIN SAVED FROM SEA BY PORPOISES

Port Lavaca, Tex. (AP) — A shrimp boat captain, E. J. Lemaire, 55, says he owes his life to friendly porpoises that helped him after he fell overboard last Tuesday.

Lemaire fell into the Gulf of Mexico five miles from the nearest land, Matagorda Island, and no one on the craft noticed.

He began swimming and soon found himself accompanied by a playful school of porpoises that nudged him, pushed him and nearly carried him to shore, Lemaire said.

Finally rescued from the island, he said he would never have made it had it not been for the porpoises. He is recovering from shock and dehydration.

Other fishermen said the waters were infested by sharks and that the porpoises kept the maneaters away as Lemaire headed for land.

(9)*

LUNG CANCER AND WOMEN SMOKERS

Associated Press

Washington

Death rates from lung cancer and other diseases are higher among women who don't smoke, according to new findings issued by the American Cancer Society last night.

However, the society reported, total death rates are considerably higher among men than among women.

Releasing a report on a survey of the smoking habits of 1,003,229 subjects, including 440,558 men and 562,671 women, the society said: "This is the first time a very large group of women has been traced for the effects of cigarette smoking. Previous reports (by the society) have dealt with the effects of cigarette smoking on disease and death rates among men."

Since the survey began in late 1959, there have been 43,221 listed deaths, male and female. The subjects ranged in age from 35 to 84 when the study started.

*critical clipping

(10)

REVOLUTION INDOORS

BOSTON MUSEUM HOUSES TREASURES OF FIGHT FOR INDEPENDENCE

Boston, Mass.—Many of the most famous and interesting mementos of Revolutionary days are housed under one roof in Boston's Museum of Fine Arts. Portraits of colonial heroes, silver by Paul Revere and fine examples of furniture from the days of the pilgrims are only a few of the treasures to be seen.

Perhaps the best known painting is Gilbert Stuart's portrait of George Washington, which was never finished. Once the artist captured the likeness of Washington's face, he stopped work on the painting and used it as a model for dozens of copies. Even Washington himself had to settle for a copy. The portrait appears on the dollar bill.

The museum honors other colonial leaders in a hall of portraits. The visitor sees Samuel Adams, father of the Revolution, John Hancock, first signer of the Declaration of Independence, Alexander Hamilton, Benjamin Franklin, Daniel Webster and a dozen others.

Historic Ale Bowl

The museum has the largest and finest collection of Revere silver in the country, including its most prized possession, the Revere Liberty bowl. The silver bowl was commissioned by the Sons of Liberty, a secret organization of patriots. Legend has it the bowl was filled with ale and passed around at secret meetings at the Liberty Tree tavern—a favorite gathering place for the patriots.

Hanging near the Revere bowl is a vigorous portrait of Paul Revere by John Singleton Copley. In the painting Revere holds a teapot which can now be seen in its case directly beneath.

Furniture of Period

More than a dozen period rooms from Europe and New England have been installed in the museum—faithfully preserving original paneling, mantels, wallpaper and furniture.

The Samuel MacIntire rooms of 1801 show the gracious interior of a house designed for a sea captain's daughter. MacIntire did much of the carving himself.

Everyday scenes of early life in America are shown in primitive landscapes and portraits of farmers and tradesmen by unknown artists. Silhouette paintings of children decorate small drawing rooms; other rooms have framed samplers or embroidered pictures of fanciful outdoor scenes.

(11)

SURVEY ON FOOD PREFERENCES

By Patricia McCormack
United Press

New York

When the bride and bridegroom say "I do" to all the wedding day vows, does it mean that she will automatically like all the foods he likes—and the reverse?

It would seem so, according to sociologists who tend to believe that marriage has a stabilizing effect on people's lives.

The results of a survey of 1000 housewives nationwide show that where food's concerned, many changes take place in preferences of the pre-wedding versus post-wedding days.

First off, the women were asked if changes had occurred in their eating habits—and those of their spouses. Sixty-four percent said yes.

Reasons

Among the reasons offered most often for changes in the palate were the following:

Family had moved to a part of the country where foods they had never before tried were popular fare. They liked them and the switch was on.

Husband landed a job that required him to attend business luncheons. He developed a liking for new dishes, came home and ordered same—and the "I do" gal of the wedding day smilingly filled the order.

An illness or physical problem dictated a change in original food pleasures. The family "fell to" and adjusted tastes.

College Attitude Survey

Name________________________ Age_______Male ___________________Female__________

This is part of a study concerning what college students think about a number of questions. The best answer to each question is YOUR PERSONAL OPINION. Of course, there are no right or wrong anwers. We have tried to cover a few different points of view. You may find yourself agreeing strongly with some of these statements, disagreeing just as strongly with others (and perhaps uncertain about others). Use the following code to indicate your agreement or disagreement with each statement:

AS—Agree Strongly, A—Agree, N—Neutral, D—Disagree, DS—Disagree Strongly

___A college education should concentrate more on preparing an individual for a specific vocation rather than on general liberal arts.

___Fraternity membership usually results in severe limitations on originality and productivity.

___Women have the right to compete with men in every sphere of economic activity.

___Capital punishment definitely helps to prevent crime.

___Schools should be allowed to have nondenominational prayers as part of their program.

___Women should not be permitted to hold political offices that involve great responsibility.

___Capital punishment is not necessary in modern civilization.

___Divorce is desirable for adjusting errors in marriage.

___We should have military training in our schools.

___Girl children should be made to feel that they are fully equal to the boys in the family.

___Fraternities help to mold the college man into a well adjusted and productive member of the society.

___Capital punishment is not an effective means of crime prevention.

___A person should have the right to marry and divorce as often as he chooses.

___Adequate preparation for war is the best guarantee of international peace.

___Fraternities do far more harm than good.

Readership Survey (Recall)

List below all of the articles that you remember reading. Use only a word or phrase to identify the article. For example, if there was an article about a debate on the recognition of Red China, you could write "Red China," "China debate," etc. After you have identified the article, indicate to what extent you read the article by circling one or more appropriate words in the right-hand column.

Head—just read headline, but did not read any of the story.
Lead—read the headline and the first one or two lead sentences.
Skim—skimmed entire story.
Read—read entire story.
Care—read entire story *carefully* or *twice*.

1. ______________________________ HEAD LEAD SKIM READ CARE

2. ______________________________ HEAD LEAD SKIM READ CARE

3. ______________________________ HEAD LEAD SKIM READ CARE

4. ______________________________ HEAD LEAD SKIM READ CARE

5. ______________________________ HEAD LEAD SKIM READ CARE

6. ______________________________ HEAD LEAD SKIM READ CARE

7. ______________________________ HEAD LEAD SKIM READ CARE

8. ______________________________ HEAD LEAD SKIM READ CARE

9. ______________________________ HEAD LEAD SKIM READ CARE

10. ______________________________ HEAD LEAD SKIM READ CARE

11. ______________________________ HEAD LEAD SKIM READ CARE

Readership Survey (Recognition)

You may have read some of the articles identified below. Circle the appropriate word after each article as you did on the preceding questionnaire. If you do not remember seeing an article, circle DID NOT SEE. If you have already mentioned noticing the article on the preceding questionnaire, cirlcle ALREADY NOTED.

1. "Survey on Food Preferences" (marriage changes eating habits)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
2. "Revolution Indoors" (museum holds treasures of fight for independence)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
3. "Policewomen Ask Same Pay as Patrolmen" (equal pay for equal work)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
4. "Best Coin Protection" (collectors guard against tarnish)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
5. "Nation's First Married Deacon Hardnosed About Military Life" (against military rebellion)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
6. "Dramatic 3-Alarmer in Mission District" (23 elderly rescued)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
7. "Marijuana's Harm Proof is Lacking" (no scientific proof)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
8. " 'Macabre' Jobs for Orphans" (take part in funeral services)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
9. "Lung Cancer and Women Smokers" (Cancer Society releases new findings)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
10. "French Plan For Fallout" (atomic test in Pacific)	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE
11. "Equador Frees Four U.S. Boats" (dispute territorial water limits)	HEAD LEAD SKIM ALREADY NOTES	READ CARE DID NOT SEE
12. "Syria Coup—Hafez is Overthrown" (political crisis in Damascus	HEAD LEAD SKIM ALREADY NOTED	READ CARE DID NOT SEE

Comprehension Questionnaire

The questions on this page ask you to describe one of the articles, *"Lung Cancer and Women Smokers,"* in more detail. If you did not read this story or do not remember the details, check here ☐ and go on to the next questionnaire.

1. What new findings were released in a report by the American Cancer Society (one or two sentences)?

2. What does the new report indicate about death rates from lung cancer and other diseases (circle one letter only)?

   (a) Death rates are *HIGHER* for women who *DO* smoke. (*LOWER* for women who *DO NOT* smoke.)
   (b) Death rates are *HIGHER* for women who *DO NOT* smoke. (*LOWER* for women who *DO* smoke.)
   (c) There are *no really significant differences* between death rates for women smokers and nonsmokers.
   (d) I do not remember what the article stated about the relationship between death rates from lung cancer and smoking in women.
   (e) The article did not specifically mention the relationship between death rates from lung cancer and smoking for women.
   (f) Other (explain) ______________________________

3. *If* you thought that there *was* a relationship between smoking and lung cancer stated in the article (2a or 2b above)

   (a) Do you personally believe it? (circle a number)

Definitely believe it	1	2	3	4	5	6	7	8	9	Definitely do *not* believe it

   (b) Whether or not you believe it, what do you think accounts for the relationship (if there was one mentioned)?

4. Other comments or feelings about the story:

Smoking Questionnaire

1. At the present time are you a cigarette (check one):

SMOKER?	Yes, but want to quit	☐
	Do *not* want to quit	☐
NONSMOKER?	Never smoked	☐
	Quit	☐

2. If you are a SMOKER
   (a) Approximate number of cigarettes per day ________
   (b) How many times have you tried to quit? ________
   (c) Have you ever tried to cut down or switch from one form of tobacco to another? ________

3. If you are a QUITTER
   (a) How long ago did you quit ________ yrs?________ mos.
   (b) Previous number of cigarettes per day ________

4. Answer the following whether or not you currently smoke:
   (a) I believe that smoking can cause cancer:
   agree strongly 1 2 3 4 5 6 7 8 9 disagree strongly
   (b) I believe smoking can cause heart disease:
   agree strongly 1 2 3 4 5 6 7 8 9 disagree strongly
   (c) I dislike smoking intensely 1 2 3 4 5 6 7 8 9 I thoroughly enjoy smoking

Recall and Recognition Data Sheet
Critical Article (No. 9)

	SMOKERS			NONSMOKERS		
	Satisfied	Want to quit	Total	Never Smoked	Quitters	Total
1. Number of Subjects (*N*)						
2. (a). *Recall*						
(b). *Recognition* but not Recall						
(c). *Neither* Recall nor Recognition						
% Recall a/N						
% Recognition $(a+b)/N$						
3. For those subjects responding 2a or 2b above: Head						
Lead						
Skim						
Read						
Care						

Recall and Recognition Data Sheet
Control Article (No.)

	SMOKERS			NONSMOKERS		
	Satisfied	Want to quit	Total	Never Smoked	Quitters	Total
1. Number of Subjects (*N*)						
2. (a). *Recall*						
(b). *Recognition* but not Recall						
(c). *Neither* Recall nor Recognition						
% Recall a/N						
% Recognition $(a+b)/N$						
3. For those subjects responding 2a or 2b above: Head						
Lead						
Skim						
Read						
Care						

Recall and Recognition Data Sheet
Control Article (No.)

	SMOKERS			NONSMOKERS		
	Satisfied	Want to quit	Total	Never Smoked	Quitters	Total
1. Number of Subjects (*N*)						
2. (a). *Recall*						
(b). *Recognition* but not Recall						
(c). *Neither* Recall nor Recognition						
% Recall a/N						
% Recognition $(a+b)/N$						
3. For those subjects responding 2a or 2b above: Head						
Lead						
Skim						
Read						
Care						

Comprehension Data Sheet
Critical Article (No. 9)

	SMOKERS			NONSMOKERS		
	Satisfied	Want to quit	Total	Never Smoked	Quitters	Total
1. Number of Subjects (*N*)						
2. Distortion (a)						
Correct (b)						
Incorrect (c, d or e)						
% Distortion *a/N*						
% Correct *b/N*						
3. For those subjects responding 2a or 2b above: Degree of Belief (3a)						

Selective Exposure
Notes for Experimenter

(Be sure to read complete description first.)

1. Arrange articles in room with minimum distractions with No. 1 on the left and the critical item 9th from the left.
2. Recruit a subject "to help you standardize some very brief tests you are working on."
3. Bring Subject (*S*) to table and move some article aside to make room for *S* to work on the filler questionnaire (page 265).
4. Say "Here is the first of two very brief attitude tests," give *S* the filler test (page 265).
5. Leave room. Carefully time absence. (See complete description.)
6. Remove articles from view when you return.
7. Say "Now I would like you to fill out this recall questionnaire (page 267). It asks you about the articles that were on the table because I am actually interested in the kinds of information people notice, read, and remember.
8. Take up recall questionnaire and pass out recognition questionnaire (page 269).
9. When the *S* is finished, take his questionnaire and look at it. For *S*s who indicated that they recalled or recognized the critical article on smoking, give them the comprehension questionnaire (page 271).
10. Take up comprehension questionnaire and pass out smoking questionnaire (page 273).

Research Example

Preference for Dissonant Information[1]

JONATHAN L. FREEDMAN

Ss heard a tape-recorded interview which was designed to make the interviewee sound either very well qualified for an international conference or very poorly qualified. Ss rated the interviewee and were then given a choice of reading an evaluation of him that agreed with their rating or one that disagreed with it. Of 18 Ss, 17 chose the evaluation that disagreed with their rating. A number of possible explanations of this result are offered, and supplementary data are presented to support them.

If people are given a choice of exposing themselves to consonant or dissonant information, which will they prefer? A number of studies have investigated this problem (Adams, 1961; Brodbeck, 1956; Mills, Aronson, & Robinson, 1959; Rosen, 1961; Sears & Freedman, 1963, 1965), and this work has been generally interpreted as demonstrating a preference for consonant information. As Steiner (1962) recently pointed out, however, none of these studies has produced unequivocal results. The weakness and ambiguity in the findings suggest that subjects' preference may depend to a large extent on the particular experimental situation employed. This dependence appears to be well illustrated in the present study, which was designed as one more attempt to demonstrate the expected preference for consonant information but which, as will be seen, produced very strong results in the opposite direction. Since the explanation of this finding which will be offered is admittedly post hoc, it is more appropriate to present the experiment first and the explanation later rather than attempting to suggest that the results were expected.

SOURCE. *Journal of Personality and Social Psychology* 1965, **2**, (2), 287–289.

[1]This study was supported, in part, by Grant GS-196 from the National Science Foundation. The author is grateful to John Steinbruner for assisting in the running of the experiment and to David Sears with whom many of the ideas and problems in this paper have been discussed.

Method

The general design was to have subjects make a decision and then give them a choice between reading a consonant or dissonant communication. The subjects

were 18 students, 9 men and 9 women, in an introductory education course at Stanford University. They were run individually. The instructions informed the subject that the study was an investigation of person perception and interpersonal judgments, and the importance of these abilities was explained in detail. It was mentioned that the ability to evaluate others correctly was a critical factor in career success and that it correlated very highly with general intelligence. The subject was told that he would hear an interview between a student who was applying to participate in an overseas conference and the person in charge of the conference, and that it was the subject's task to evaluate the student after hearing the interview.

The subject heard either an interview designed to make the student sound very good or one designed to make him sound very bad. The interview was prerecorded on magnetic tape and lasted approximately 6 minutes.

After the interview, the experimenter told the subject to think carefully about what he had heard and then to record his judgments on an evaluation form. The major question on this form was: "Should the applicant be accepted to participate in the conference?" which was answered by placing a mark on an 8-point scale labeled "definitely no" at one end, and "definitely yes" at the other. The subject then rated the applicant on nine scales similar to the one just described except that the end points were labeled with contrasting adjectives such as unfriendly-friendly, intelligent-stupid, etc. These additional rating scales were included to make it seem more plausible to the subject that we were studying interpersonal perception and that this was the only interview he would have to rate.

When the evaluation form was completed, the experimenter said that we would now like the subject to read some evaluations of the candidate that had been written by others, that we had two such evaluations written by people who knew him well, and that he could read either of these. The experimenter also explained that one of these evaluations was quite favorable, while the other was quite unfavorable. The subject was then asked which he would like to read. His choice was recorded, and the experimenter questioned the subject for some time concerning why he had made that choice. The experiment was terminated, the deception explained, and the subject was asked not to tell others about the study.

Results

The eight subjects who heard the unfavorable interview all rated the candidate unfavorably, with a mean rating of .75 on a scale ranging from 0 to 8. Nine of the 10 subjects who heard the favorable interview rated the candidate favorably, with a mean of 5.94. None of these 17 gave a rating that fell within 1 point of the middle of the scale. That is, they were all quite extreme. One subject who heard the favorable interview gave the candidate a rating of 1.5, a very unfavorable rating. The results will be presented separately for this subject.

The major results are the preferences

shown by the subjects between the consonant or dissonant information. All 17 subjects chose to read the dissonant communication. Subjects who had rated the candidate favorably chose the unfavorable evaluation; those who had rated him unfavorably, chose the favorable evaluation. Although statistics on such a unanimous choice seem superfluous, it is a highly significant effect ($p < .001$)[2] from binomial distribution).

The one subject who heard the "favorable" interview and rated the candidate unfavorably chose to read the unfavorable evaluation. This was a choice of information consonant with her decision but dissonant with the interview. If it is considered a choice of consonant information, however, it does not change the total result appreciably (p for a 17 and 1 split is still $< .001$)

Thus, subjects in this situation strongly prefer nonsupportive to supportive information. This is not simply a lack of preference such as has often been reported in the past (e.g., Brodbeck's 1956, overall finding; Feather, 1963; the negative groups in Mills et al., 1959), but a clear preference for dissonant information. How may this result be explained?

One explanation is simply that the dissonant information was seen as more useful and was, therefore, preferred. There is considerable evidence that the more useful information is, the more subjects seek it out (Canon, 1964; Freedman, 1965; Mills et al., 1959). Although subjects in the present study were not told to expect more than one interview and, from the amount of time remaining, it must have been clear that additional interviews were unlikely, it is possible that they may still have expected to have to evaluate other candidates. Under these circumstances it is possible that the dissonant evaluation (i.e., the evaluation that disagreed with the subject) was expected to contain more new information than the consonant evaluation, and the former was, consequently, expected to be more useful in improving performance on the future decisions. During the question period after the choice, only four of the subjects choosing the dissonant communication specifically mentioned usefulness, utility, or a similar factor in explaining their choice, but this does indicate that at least for some subjects information utility was an important consideration.

Another factor which may have influenced the results was the perceived interest of the information. Subjects had just gotten a very one-sided picture of the candidate. An evaluation which disagreed with this picture would presumably present a new outlook, possibly new facts, certainly a new opinion. Other things being equal, this dissonant evaluation would probably be more interesting than one which agreed with both the picture presented in the interview and the subject's own opinion. If the dissonant evaluation were seen as potentially more interesting, subjects would probably tend to prefer it to the consonant evaluation. This possibility is given some support by subjects' responses during the question period. Subjects often said that they were "curious about the other side of the picture" or that

[2] Notations such as $p < .001$ are explained in the Introduction, pp. 15-16.

they "wanted to know what someone else thought about this guy." Of the 17 subjects choosing to read the dissonant evaluation, 8 made statements similar to these.

Although both usefulness and interest of the information probably contributed to the preference for dissonant information, the size of the effect (as compared with those generally reported in related studies) suggests that additional factors may be operating. Freedman and Sears (1963) have offered a rather complex analysis of the present problem which appears to be consistent with the results reported here. The analysis involves the subject's confidence in his position and the extent to which he perceives the issue as controversial. In most studies on selective exposure, the subject takes a position on an issue and is then asked to rate articles which either agree or disagree with that position. This means that he is presented with the titles of articles which supposedly disagree with him. If the subject did not previously know that the issue was controversial, this in itself should arouse dissonance because it essentially tells the subject that a dissonant article exists. Under these circumstances the question is not one of avoiding dissonance, but of reducing dissonance that is already present.

At this point confidence may be a major determinant of the subject's strategy. If he is very confident that he is correct, it may well be that the best course of action is to seek out the dissonant article and attempt to refute it. He is confident that he can accomplish this, and he may prefer to try it rather than not looking at the article and being uncertain how strong the counterarguments are. If the subject has little confidence, however, it should generally be safer for him to avoid the dissonant article and attempt to reduce the dissonance by internal processes. Exposure to the article will probably increase dissonance even more since he will be unable to refute the arguments. Thus, with issues not previously known to be controversial, high-confident subjects should seek out dissonant information; whereas, low-confident subjects should prefer consonant information.

If, on the other hand, the subject already knows that the issue is highly controversial, the knowledge that a dissonant article exists will arouse little or no dissonance because the subject already knows that such articles must exist. Under these circumstances, it is a question of avoiding dissonance and all subjects should tend to prefer consonant articles.

The present situation fits rather well into the low-controversial, high-confidence cell of this analysis. The interviews were quite one-sided and presented a candidate who was either obviously very good or obviously very bad. There was evidently some room for doubt since one subject did disagree with the others' judgment; but all other subjects gave very similar ratings to their respective candidates. Thus, it is reasonable to suppose that subjects saw the decision as rather clear-cut with little room for error or disputation. In addition, they were probably quite confident of their decision. When they were told that a dissonant evaluation existed, dissonance should have been aroused and they sought out the dissonant evaluation in an attempt to refute it.

Over half of the subjects choosing the dissonant evaluation said, in essence,

that they were confident of their own decision and could not believe that the other was correct. As opposed to this, the one subject who rated the candidate inappropriately and who was probably less confident of her decision than were the other subjects, chose the consonant evaluation and said that she wanted to see if anyone really agreed with her point of view. Thus, this analysis seems to fit the results fairly well and to be supported by statements made by the subjects. It is, however, purely speculative at this stage, and an experiment designed to test it explicitly is needed before it can be adequately assessed.

The one point which should be clear from this study is that preference between dissonant and consonant information is considerably more complicated than it has generally been pictured. Many factors probably affect it, and at least some of them tend to make subjects prefer dissonant information. The present study presents strong evidence that under some circumstances subjects will choose dissonant over consonant information. They may be seeking out the dissonant information only because they hope to refute it, but the fact remains that their preference is for information that is clearly labeled as dissonant.

References

Adams, J. S. Reduction of cognitive dissonance by seeking consonant information. *Journal of Abnormal and Social Psychology*, 1961, **62**, 74–78.

Brodbeck, May. The role of small groups in mediating the effects of propaganda. *Journal of Abnormal and Social Psychology*, 1956, **52**, 166–170.

Canon, L. K. Self-confidence and selective exposure to information. In L. Festinger (Ed.), *Conflict, decision, and dissonance.* Stanford: Stanford Univer. Press. 1964, Pp. 83–96.

Feather, N. T. Cognitive dissonance, sensitivity, and evaluation. *Journal of Abnormal and Social Psychology* 1963, **66**, 157–163.

Freedman, J. L. Confidence, utility, and selective exposure: A partial replication. *Journal of Personality and Social Psychology,* 1965, in press.

Freedman, J. L., & Sears, D. O. Voters' preferences among types of information. *American Psychologist,* 1963, **18**, 375 (Abstract)

Mills, J., Aronson, E., & Robinson, H. Selectivity in exposure to information. *Journal of Abnormal and Social Psychology,* 1959, **59**, 250–253.

Rosen, S. Postdecision affinity for incompatible information. *Journal of Abnormal and Social Psychology,* 1961, **63**, 188–190.

Sears, D., & Freedman, J. Commitment, information utility, and selective exposure. Report No. 12, 1963, Office of Naval Research.

Sears, D., & Freedman, J. Effects of expected familiarity with arguments upon opinion change and selective exposure. *Journal of Personality and Social Psychology*, 1965, in press.

Steiner, I. D. Receptivity to supportive versus nonsupportive communications. *Journal of Abnormal and Social Psychology,* 1962, **65**, 266–267.

Exercise 11

Group Behavior

Individuals working together in a group seldom operate in parallel; that is, each of the group members do not have to perform all of the tasks necessary for the group to accomplish its objectives. Instead, each member can do what he knows how to do well and his abilities can complement instead of duplicate those of his fellow group members.

In structured groups, roles or functions are formally assigned. One person may be designated as chief or boss, another as clerk or secretary, while a third may serve as a technical expert. But even in less formal settings, such as rap sessions or coffee breaks, different roles for the different group members naturally develop early in the evolution of the group. Sometimes these roles are obvious to an outside observer; for example, it is usually easy to identify a man who assumes the function or role of task-leader; he rarely tells jokes, and he tries to keep the group "on the track" toward completing the task. The leader asks for evidence to support opinions, he is critical of bad ideas, and he is usually the person to suggest that it is time for the group to move on to a new topic. However, many roles are more subtle and hence more difficult to identify. For example, Slater (1955) has suggested that the leadership role may be further differentiated: there may be a task leader (the person rated as having the best ideas and guidance ability), and a socio-emotional leader (the best-liked group member who soothes hurt feelings and reduces tensions).

In this exercise you will either experience or observe the emergence of different role behaviors within a small group. The instructor may choose either to divide the entire class into groups of five or more people (option 1), or to assign four class members to serve as a demonstration group while the

remainder of the class observes and records their behavior (option 2). If the instructor chooses option 1 (everyone participating), you are to *read only your own role* before beginning the group discussion. After the discussion, your task will be to guess the roles of the other group members. Each group member should be assigned a letter designation; A, B, C, D or E. Do not inform the other group members about the part you will play and do not read the role of the other group members. Tear out only the page labeled with your own letter and read the role description and requirements.

If option 2 (demonstration group) is chosen, the four participants will be assigned one role, A, B, D, or E. *All nonparticipating class members will tear out and read role C only.* For either option, larger groups can be formed by assigning more than one person to the same role, or the instructor may wish to create descriptions for additional roles.

When everyone has read their role, start a five to ten minute discussion on a controversial or interesting topic, for example, "Should the draft be abolished in favor of a voluntary army?" or "What difference would it make if man were equipped with four fingers but no thumb?" After the discussion is completed, each group member should fill out the questionnaire on page 311 and continue reading on page 302.

Role A

You have a leadership role. Make sure that everyone talks and contributes to the discussion. Try to mediate arguments. Occasionally you should summarize the progress of the group, and remind members of points that have previously been made. In other words you should act as chairman of the group. You are responsible for the group accomplishing as much as it can in an orderly manner.

Role B

Your job is to keep the group members happy and satisfied with the group discussion. Try to prevent conflicts and disagreements. Keep the tension low by suggesting compromises and being cheerful or humorous. Agree with what other people say and make them feel that they are making a constructive contribution to the group.

Role C

You may make talk if you want; but your main job is secretarial. You will keep a record of all group communications on the *Group Interaction Tally* provided (page 317). Although it will be difficult to categorize the statements, it will be fairly easy to keep a tally of the number of times each member speaks. If a long speech contains several different thought units, you may use a tally mark for each separate thought unit in the speech. Use only one category for each unit of speech. Other categories may be added in the space at the bottom of the Tally if desired.

Role D

You have the role of a critic. Make sure that each group member states his ideas clearly and backs them up with evidence. Be sure that the group never accepts an idea that isn't sound.

Role E (optional)

You may make contributions to the discussion; however, your main job will be to record the ideas that the group produces. Record as many ideas as you can even if they seem inappropriate or stupid. Note who initiates each idea and how it is received.

DISCUSSION

Do not read this section until you have answered the questionnaire on page 311 (and are instructed by your teacher to continue reading).

First read the roles played by other participants. How accurate were your guesses? Did your guesses agree with those of the other observers? The observers may agree on what role a group member played; but they may not think he played the role described in the instructions. What does that mean? Did the "role player" fail to correctly portray his intended role?

Discuss the kinds of behavior that you used to make your judgments as an observer. For example, what kind of behaviors would lead an observer to think a certain group member was a leader?

Did two people compete to occupy the same role? Was there more than one leader? Each participant was instructed to act according to a specific description of his role, but did personality differences among the participants affect their role behavior?

Some social psychologists have argued that anyone who is in the right place at the right time will be perceived as a leader. For example, Leavitt (1951) has shown that the central man in a communication network is invariably perceived as the leader—regardless of his personal characteristics.

Discuss the relationship between situational and personality factors in leadership in your discussion group. Did any "nonleader" *personality* types play leadership *roles* in this exercise? Each participant can discuss difficulties created by inconsistencies between his personality or life style and his assigned role.

DATA ANALYSIS

Research has shown that the man who talks the most is often perceived as the leader, at least in groups that do not meet over a long period of time (Bavelas et. al., 1965). Use Table 1, page 323, to find out whether this was true for your group. A measure of actual participation can be obtained from the observations made by the man playing role C. In the columns labeled "Actual Participation" in Table 1, enter the number of actual statements made by each participant and then rank each participant starting with the rank 1 for the person who actually talked most.

Next, record "Perceived Participation" in Table 1. To accomplish this, average the ranks assigned by each observer on the questionnaire item "Who talked the most?" If there were five group members, each member's perceived participation would be obtained from averaging the ranks assigned by the remaining four group members. For example, if member A was ranked highest participator by B and C, and second highest by D and E, his average

perceived participation would equal 1.5 $[(1 + 1 + 2 + 2)/4 = 1.5]$. Rank perceived participation assigning a rank of 1 to the man receiving the highest average ratings.

In a similar manner, record perceived leadership ratings in the third columns using the questionnaire item "Who acted most like the group leader?" Ranks should be assigned as before. If desired, similar rankings can be made for the other questionnaire ratings, or combined ranking can be made. For example, you may prefer to sum ratings for the questionnaire items dealing with best ideas, guidance, and leadership as an overall index of perceived leadership. This procedure would yield index numbers for each participant from each observer.

When the rankings are completed you are ready to determine how closely talking (actual and perceived) are related to perceived leadership. To determine this relationship you could calculate a rank-order correlation (Appendix C) between columns 1 and 3, and 2 and 3 in Table 1. However, since you have only five pairs of numbers, you can simply inspect the data to determine the degree of relationship. In many cases you will note that the rank order of group members for leadership and participation are identical.

What are some of the reasons that the leader is usually the biggest talker? Remember that this is a correlational finding and you may speculate about the direction of causality (Introduction, page 11). It is plausible that talking leads to perceptions of leadership, but it is also possible that good ideas and other "leadership acts" are rewarded by other group members and that this reinforcement leads to more talking.

Discussion question: What differences might be found between permanent and short-lived groups?

OPTIONAL GROUP BEHAVIOR ANALYSES

Following are three examples of the many additional possibilities for using the kind of data gathered in this exercise to study group behavior.

Trends in Group Behavior Over Time

The easiest way to study changes in group behavior over time is to have one demonstration group (one member playing each role) discuss several topics sequentially. The class or the instructor can prepare three or four discussion topics each of which is to be discussed for a fixed time period, say 5 minutes. At the end of each time period, sociometric questionnaires (page 311ff) are completed and totaled. A new Group Interaction Tally (page 317ff) is started for each new discussion topic. Participation and other ratings for each group member can be plotted over the discussion periods. An example

is presented in Fig. 1 for changes in amount of talking over three discussion periods. What trend is demonstrated in the example?

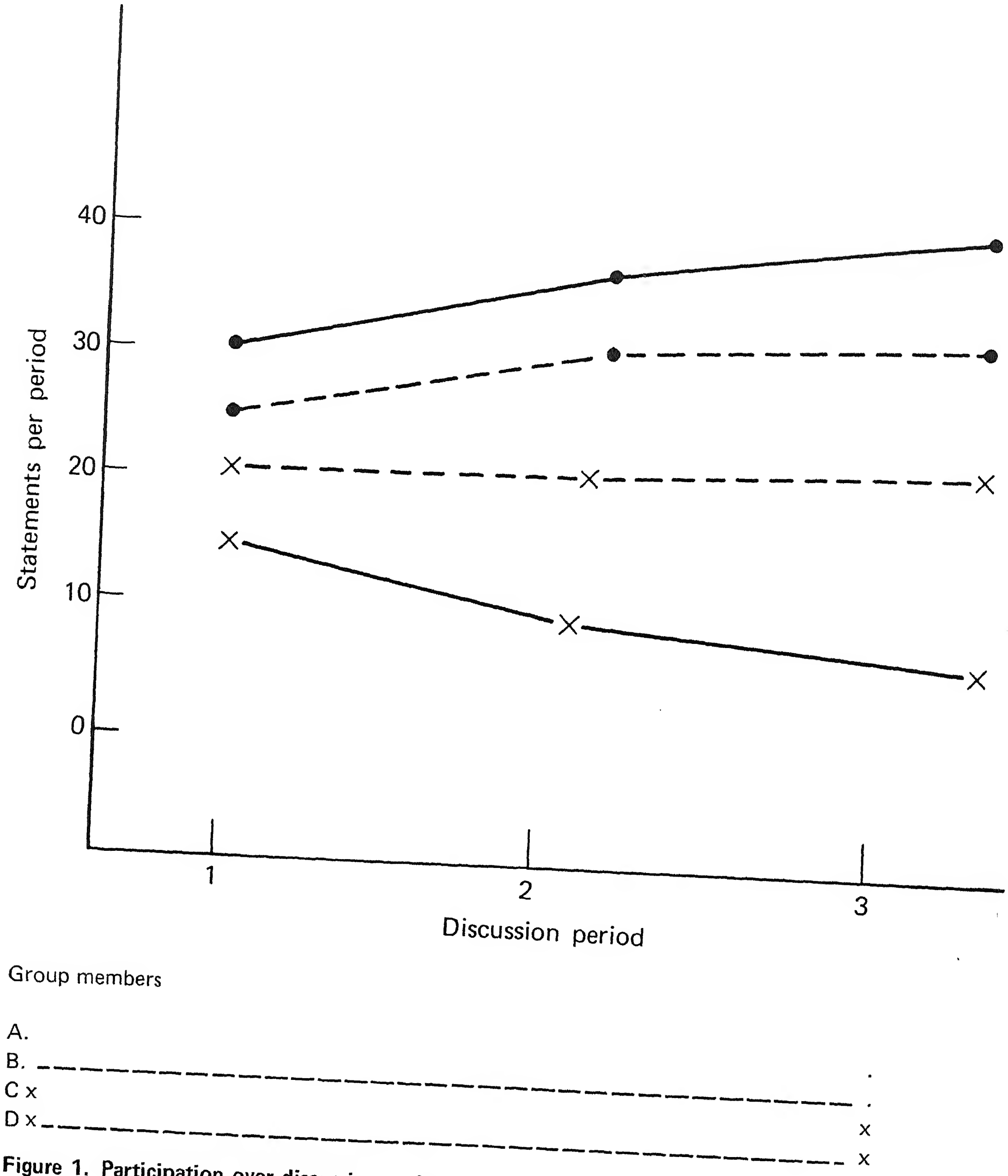


Figure 1. Participation over discussion periods.

The most obvious interpretation for the trend in Fig. 1 is that the changes are the result of practice, growth, or motivation. Several theorists (for example, Bales, 1950; also see the Research Example) have suggested that all new groups go through definite phases as the members get to know each other. But we cannot rule out one possible confound or methodological artifact. A comparison between discussion period 1 and discussion period 2 *is also a comparison between topic 1 and topic 2.* Topic and time period are confounded; either is a plausible cause. To eliminate this confound you would have to run several demonstration discussions with the topics discussed in at least several possible orders.

For a simple example with only two topics and two discussion periods, see Fig. 2. If subjects' behaviors are different in the left column than in the right

		Topic 1	Topic 2
Discussion Period	First		
	Second		

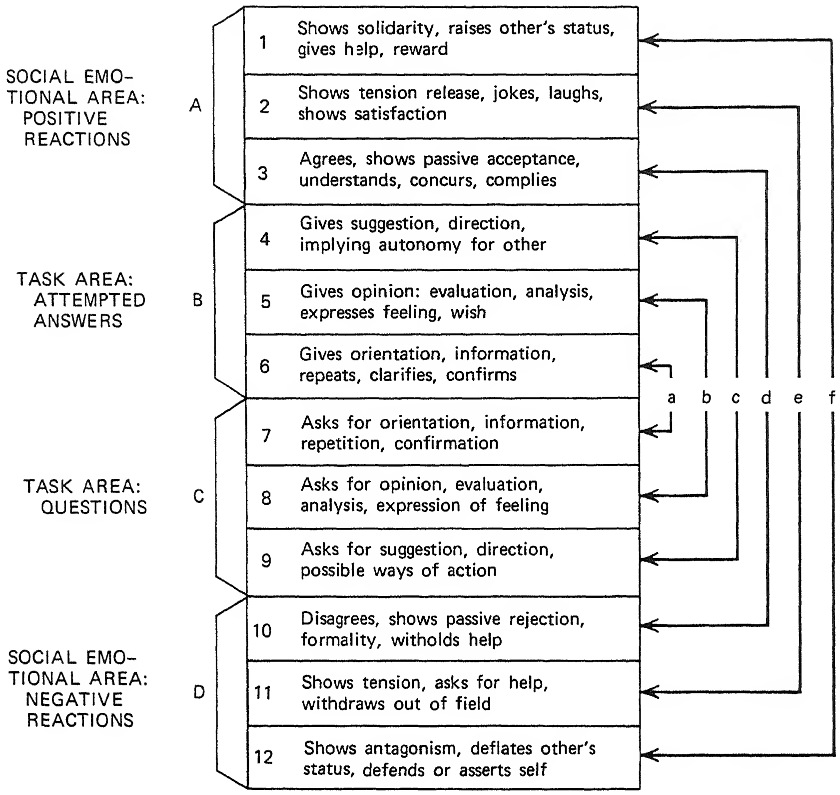
Figure 2. Confounding of topic and discussion period.

column, then we know that the difference is caused by the topic because half of the subjects discussed topic 1 first and half discussed it second. Similarly, if behaviors are different from period 1 to period 2 (if the top row is different from the bottom row), then the pattern of behavior is at least partly caused by the changes over time, regardless of topic.

It is important to keep this possible confound in mind, but in this exercise you can ignore the possible effect of topic because you will simply observe one demonstration group discuss several topics.

Developing Rating Categories

The number of categories used for rating group discussions can be increased or decreased. The number of categories certainly should not exceed the number that the observers can reliably rate. A very popular system for recording group interaction processes that has been developed by Bales (1950) is presented in Fig. 3. Some data yielded by this system are shown in Fig. 1 of the Research Example, p. 327.



F igure 3.

In addition to expanding or collapsing categories, it is possible to develop a system that notes to whom a particular statement is addressed (either to another individual or to the entire group.) This data, which can be gathered from observing the direction of the speaker's eyes and head, makes it possible to analyze the pattern of communication in the group and may be used in answering many interesting questions dealing with social interaction. For example, does the perceived leader address more of his statements to the group at large than he does to single individuals? Does the quietest member direct all of his statements to the leader?

Reliability

One way to examine observer reliability or interjudge agreement is to look at the degree of concordance between two or more people who are observing the same behavior. If most of your class acted as observers, or if any groups had more than one observer (role C), you can check on the reliability of the observations. If there was only one observer for each group, you will not be able to do this part of the exercise.

It is fairly easy to calculate reliability for a simple variable such as amount of talking. Each observer notes the number of statements he has recorded for each group member and then ranks each participant on a scale from most talkative to most quiet. If there are only two observers or only two that you are interested in, you can compute a rank-order correlation as illustrated in Appendix C. Alternatively, you can simply calculate the percentage of times that the two raters agree exactly on the ranks assigned to group members or, less stringently, the percentage of times they agree exactly or disagree by only one rank.

These last two suggestions yield only rough estimates of reliability. If you wish to calculate rating agreement among a group of more than two raters, you can compare all of the possible pairs of raters and average these comparisons. For example, with four raters there are six possible pairs that can be compared. These six correlations or percentage agreements can be averaged as an index of overall agreement. An easily computed statistic known as Kendall's *W* coefficient of concordance (Siegel, 1956, page 229ff) is also available to indicate reliability for an entire group of raters.

Problems of obtaining an accurate index of reliability increase when we are interested in observer agreement on coding the meaning of statements instead of simply counting the number of statements. One index of coding agreement is defined as the number of times two raters agree on the category to which a particular statement should be assigned, divided by the total number of statements coded.

$$\text{Reliability \%} = \frac{\text{Agreements}}{\text{Agreements} + \text{Disagreements}}$$

A difficulty with this kind of index should become apparent as you compare ratings between coders: one rater may break a single speech into several parts and code them separately, while his paired rater may assign the same speech to a single category. Such discrepancies in the number of statements rated will reduce the reliability percentage. Causes of low reliability can be detected if the rating is done on a typed transcript of a tape-recorded group discussion. The transcript may be divided into statements that are then assigned to categories by two raters. Another possibility is to break the discussion period down into small time units (for example, 30 seconds) and instruct each rater to start a new data sheet every 30 seconds. This simplifies the problem of deciding whether both observers were rating the same communication.

The careful researcher recognizes the importance of including a check for rater reliability even though it may be difficult to calculate. If reliability between raters is very low, not much confidence can be put in the results.

The results may reflect rater quirks and idiosyncrasies instead of more generally normative (agreed-on) phenomena. Reliability and accuracy may improve when the observer is not participating in the discussion and is able to concentrate fully on recording the behavior. Although reliability is a problem of which all students should be cognizant, it is a problem that may be ignored in this simplified classroom demonstration.

References

Bales, R. F. *Interaction process analysis: A method for the study of small groups.* Cambridge, Mass.: Addison-Wesley, 1950.

The classic detailed account of the most popular method for categorizing group behavior (see Fig. 3). Includes thorough consideration of issues relating to the development of interaction in small groups.

Bavelas, A., Hastorf, A. H., Gross, A. E., and Kite, W. R. Experiments on the alteration of group structure. *Journal of Experimental Social Psychology,* 1965, **1**, 55–70 Also reprinted in D. Cartwright and A. Zander (Eds.), *Group dynamics,* New York: Harper & Row, 1968, 527–537.

This research demonstrates how changes in amount of participation can lead to positively correlated changes in leadership ratings.

Leavitt, H. J. Some effects of certain communication patterns on group performance. *The Journal of Abnormal and Social Psychology* 1951, **46**, 38–50. Also reprinted in E. E. Maccoby, T. M. Newcomb, and E. L. Hartley (Eds.), *Readings in social psychology,* New York: Holt, Rinehart and Winston, 1958, 546–563.

This article discusses the relative problem-solving effectiveness of four types of communication networks.

Siegel, S. *Nonparametric Statistics.* New York: McGraw-Hill, 1956.

A good "cook book" for doing nonparametric statistical tests.

Slater, P. E. Role differentiation in small groups. *American Sociological Review* 1955, **20** 300–310. Also reprinted in R. Bales, E. Borgatta, and P. Hare (Eds.), *Small groups* New York: Knopf, 1962, 498–515.

Slater distinguishes between the task leader and the socioemotional leader. He also discusses conditions that lead to increased specialization.

Questionnaire

1. Guess the role played by each group member. For example, who was the task leader?

Group Member *Role*

A ______________________________

B ______________________________

C ______________________________

D ______________________________

E ______________________________

F ______________________________

2. For the following items, rank each group member. For example if member "D" talked the most, enter "D" next to "1st" under item 2a.

(a) Who talked the most?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(b) Who had the best ideas?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(c) Who did the most to effectively guide the discussion?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(d) Who acted most like the group leader?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(e) Who was the most likeable group member?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

Questionnaire

1. Guess the role played by each group member. For example, who was the task leader?

Group Member *Role*

A ____________________

B ____________________

C ____________________

D ____________________

E ____________________

F ____________________

2. For the following items, rank each group member. For example if member "D" talked the most, enter "D" next to "1st" under item 2a.

(a) Who talked the most?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(b) Who had the best ideas?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(c) Who did the most to effectively guide the discussion?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(d) Who acted most like the group leader?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(e) Who was the most likeable group member?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

Questionnaire

1. Guess the role played by each group member. For example, who was the task leader?

Group Member *Role*

A ______________________________

B ______________________________

C ______________________________

D ______________________________

E ______________________________

F ______________________________

2. For the following items, rank each group member. For example if member "D" talked the most, enter "D" next to "1st" under item 2a.

(a) Who talked the most?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(b) Who had the best ideas?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(c) Who did the most to effectively guide the discussion?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(d) Who acted most like the group leader?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

(e) Who was the most likeable group member?
1st ______ 2nd ______ 3rd ______ 4th ______ 5th ______ 6th ______

Group Interaction Tally for Use with Role C Group Members (<u>Not</u> Roles)

	Person A	Person B	Person C	Person D	OPTIONAL Person E	Person F	Person G
Task-oriented *Gives* facts, opinion, suggestion							
Asks for facts, opinion, suggestion							
Social-emotional *Positive,* harmonizing, supportive							
Negative, irrelevant, or other							
Other Categories 1.							
2.							
3.							
Total number of statements							

Group Interaction Tally for Use with Role C Group Members (Not Roles)

	Person A	Person B	Person C	Person D	OPTIONAL Person E	Person F	Person G
Task-oriented *Gives* facts, opinion, suggestion							
Asks for facts, opinion, suggestion							
Social-emotional *Positive,* harmonizing, supportive							
Negative, irrelevant, or other							
Other Categories 1.							
2.							
3.							
Total number of statements							

Group Interaction Tally for Use with Role C Group Members (Not Roles)

					OPTIONAL		
	Person A	Person B	Person C	Person D	Person E	Person F	Person G
Task-oriented *Gives* facts, opinion, suggestion							
Asks for facts, opinion, suggestion							
Social-emotional *Positive,* harmonizing, supportive							
Negative, irrelevant, or other							
Other Categories 1.							
2.							
3.							
Total number of statements							

Table 1

Group Member		1 Actual Participation (number of statements from C's observations)		2 Perceived Participation		3 Perceived Leadership	
Letter	Assigned role	Number	Rank	Average Assigned by other members	Rank	Average Assigned by other members	Rank
A							
B							
C							
D							
OPTIONAL E							
OPTIONAL F							
OPTIONAL G							

Research Example

Task Roles and Social Roles in Problem-Solving Groups

ROBERT F. BALES

During the last ten years, a number of laboratories for the study of social interaction within small groups and organizations have been started in university research centers, hospitals, clinics, and military installations. The studies and experiments I shall describe were conducted in one of these laboratories, which was established in 1947 at Harvard University.

The laboratory consists of a large, well-lighted room for the group under study and an adjoining room for observers who listen and watch from behind windows with one-way vision. The subjects are told at the beginning that the room has been constructed for the special purpose of studying group discussion, that a complete sound recording will be made, and that there are observers behind the one-way mirrors. The purpose of the separation is not to deceive the subjects but to minimize interaction between them and the observing team.

SOURCE. *Readings in Social Psychology, 3rd Edit.* Maccoby, E., Newcomb, T. & Hartley, E., Eds. Holt, 1958, 437–447.

Over a number of years we have evolved a more or less standard type of group and task which has formed the setting for a number of studies. The data I shall report came from several studies, all done under essentially the same conditions, so that a description of the most recent investigation will serve in substance for the others.

Procedures

The sample which provided data for the most recent investigation consisted of 30 five-man experimental groups. Subjects were 150 Harvard freshmen who were recruited by letters sent to a random sample of the entering class which briefly described the experiment as one concerned with group problem-solving and decision-making. Volunteers were offered a dollar an hour. The groups were randomly composed. Typically the members of a group did not

know each other, nor were they introduced to each other. In effect, they were faced with the problem of getting organized as well as with the more obvious problem that was issued to them.

The more obvious problem, which we call the standard task, involved the discussion of a human-relations case, a five-page presentation of facts about a problem facing an administrator in his organization. Members were given separate identical copies of the case to read ahead of time and were told that, although each was given accurate information, we intended to leave them uncertain as to whether they each had exactly the same range of facts. The cases were collected after they had been read by the members individually, to prevent direct comparison of typed copies, although members were allowed to take notes. The task defined for each group was to assemble the information, to discuss why the people involved were behaving as they did, and to decide what should be recommended as action for the solution to the problem presented. The groups were asked to time themselves for 40 minutes and to dictate the group solution for the sound record in the final one or two minutes of the meeting.

While the group members began to organize themselves and to solve the case problem, the observers got to work in the observation room. They systematically recorded every step of the interaction, including such items as nods and frowns. Each observer had a small machine with a moving paper tape on which he wrote in code a description of every act—an act being defined essentially as a single statement, question, or gesture. Acts ordinarily occurred at the rate of 15 to 20 per minute. The recorded information on each act included identification of the person speaking and the person spoken to and classification of the act according to predetermined categories. The categories included attempts to solve either the organizational problems of the group or the task problems by the offering of information, opinions, and suggestions.

Questions and several types of positive and negative reactions completed the set of 12 categories (see Figure 1). This method is called "interaction-process analysis."[1] The categories are meant to have a general-purpose usefulness for group research and their use is not confined in any way to the laboratory conditions described here, although the best norms exist for the standard task and the group type described here.[2]

As Figure 1 shows, on the average about half (56 percent) of the acts during a group session on the standard task fall into the categories of problem-solving attempts; the remaining 44 percent are distributed among positive reactions, negative reactions, and questions. In other words, the process tends to be two-sided, with the reactions serv-

[1] Robert F. Bales, *Interaction Process Analysis: A Method for the Study of Small Groups* (Cambridge, Mass: Addison-Wesley Co., Inc., 1950).

[2] For norms, see Robert F. Bales and Edgar F. Borgatta, "Size of Group as a Factor in the Interaction Profile," in A. Paul Hare, Edgar F. Borgatta, and Robert F. Bales, *Small Groups, Studies in Social Interaction* (New York: Alfred A. Knopf, Inc., 1955), pp. 396–413.

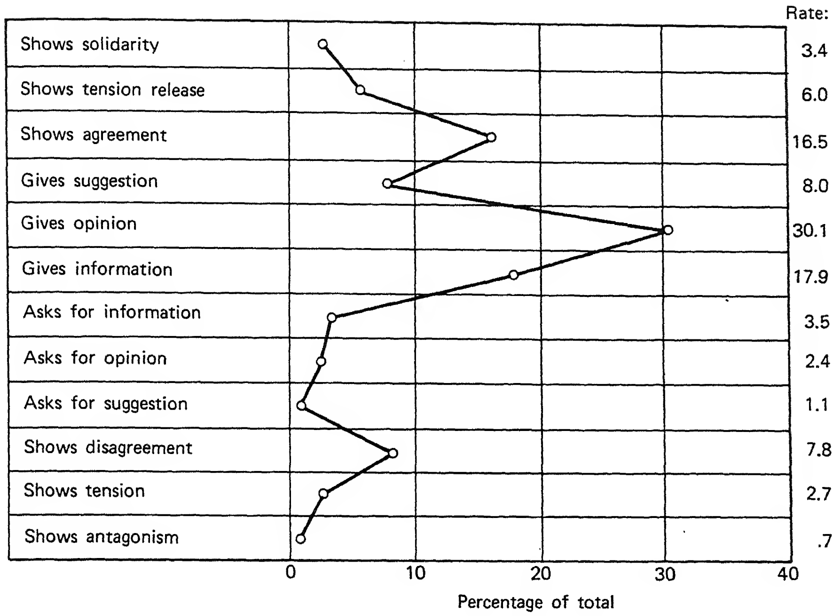


Figure 1. Types of interaction and their relative frequencies. This profile of rates is the average obtained on the standard task from 24 different groups, 4 of each size from 2 to size 7, each group meeting 4 times, making a total of 96 sessions. The raw number of scores is 71,838. (From Robert F. Bales, "How People Interact in Conferences," *Scientific American,* Vol. 192, March, 1955.)

ing as a more or less constant feedback on the acceptability of the problem-solving attempts. The following example will illustrate the pattern of interchange.

Member 1: I wonder if we have the same facts about the problem? [Asks for opinion.] Perhaps we should take some time in the beginning to find out. [Give suggestion.]
Member 2: Yes. [Agrees.] We may be able to fill some gaps in our information. [Gives opinion.] Let's go around the table and each tell what the report said in his case. [Gives suggestion.]
Member 3: Oh, let's get going. [Shows antagonism.] We've all got the same facts. [Gives opinion.]
Member 2: (Blushes) [Shows tension.]

A number of interesting generalizations can be made about the way in which rates of activity in the various categories tend to differ according to group size, time periods within a meeting, development of a group over a series of meetings, pre-established status characteristics of members, and the like.[3] The present article, however, will be

[3] For a short review, see Robert F. Bales "Some Uniformities of Behavior in Small Groups" in G. E. Swanson, T. M. Newcomb and Eugene L. Hartley (eds.), *Readings in Social Psychology* (New York: Henry Holt & Co., Inc., 1952), rev. ed., pp. 146–159.

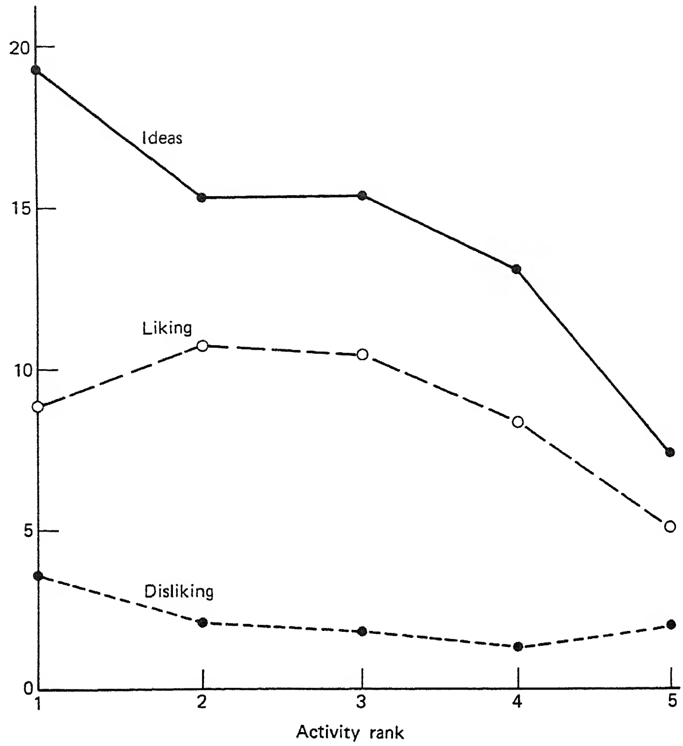


Figure 2. Average ratings* received on ideas, liking, and disliking by men of each activity rank. (From Robert F. Bales, "The Equilibrium Problem in Small Groups," Ch. IV in Parsons, Bales, and Shils (Eds.), *Working Papers in the Theory of Action,* Glencoe, Ill.: Free Press, 1953, p. 146.)

*Each entry at a given activity rank is a mean over 12 sessions for the persons who occupied that rank as of each meeting. (Four separate five-man groups were involved.) The idea index is not actually a rating, but an index obtained by adding rankings received (including self-rankings) and subtracting the total from the highest possible, 25. The like and dislike indexes are average ratings received, with the highest possible, 28.

concerned with a particular set of problems in which the interaction data have played an important part—whether there are tendencies for persons to develop different roles during interaction, even though there are no pre-established status differences, and if so, what kind, and why? There are several plausible views about this set of problems. The following account presents four distinguishable views and shows how research led from one view to another in the course of several studies.

The Hypothesis of a Single-Status Order

Perhaps the most ordinary conception of a group is that it consists of a leader and several followers who fall into a kind of status order from highest to lowest. The leader is the best-liked member of the group, participates most actively, and is felt to be the best performer of whatever task activities the group undertakes. No matter which of these criteria the researcher takes, he will come out with the same rank order of members.

The expectation that most groups are structured like this and that departures from this simple form of organization may be treated as the result of special circumstances may be called the hypothesis of a "single-status order."

This is a plausible hypothesis. It underlies much research on leadership. It is congruent with the ideological position that for good leadership it is very important that a good leader should be an all-around "great man," and it assumes that there are such men, at least relative to the other members in a given group.[4] This hypothesis assumes role differentiation but essentially only along a single quantitative dimension, leadership status.

Early in the research we began to ask group members about their likes and dislikes for each other, their opinions of who had the best ideas and who showed the most leadership, and other similar questions. We wanted to know how these questions related to each other and to our observations of interaction. The question as to whether or not there is role differentiation within a group can be reduced in part to whether group members show some consensus that certain members stand higher than others on a given criterion and whether different criteria give different status orders rather than a single-status order.

When I first began to examine data from our experimental groups, I worked under the assumption that there might be some such thing as a "simply organized group," that is, one in which the rank order of members on activity, task ability, and likeability would coincide, and that these groups would in some sense or other be the most successful or best satisfied.[5]

Figure 2 shows the results which raised a most interesting set of questions. The total interaction initiated by one man in the course of a meeting establishes the basis for ranking him relative to the others on activity. If there is a strong tendency toward a single-status order, top men on activity should also rank highest in group-member responses to such questions as "who has the best ideas," and should also receive the highest number of "liking" votes and lowest of "disliking."[6] The second man on activity should, on the average, be second highest on the other criteria of excellence, and so on. The rank order on each criterion should be highly correlated to the rank order on the other criteria.

What does Figure 2 suggest? First, there seems to be a positive correlation between activity rank and idea rank,

[4] For some evidence that there are some such men, in relative terms, see Edgar F. Borgatta, Arthur S. Couch, and Robert F. Bales, "Some Findings Relevant to the Great Man Theory of Leadership," *Am. Sociol. Rev.*, 1954, **XIX**, 755–759.

[5] Robert F. Bales, "The Equilibrium Problem in Small Groups," Ch. IV in Talcott Parsons, Robert F. Bales, and Edward A Shils (eds.), *Working Papers in the Theory of Action* (Glencoe, Ill.: Free Press, 1953.)

[6] The actual questions used are presented in the source indicated at the foot of Figure 3. They are omitted in the present paper for the sake of brevity.

although the second man seems a little low. But on liking-received rank, there is a marked discrepancy. The top man on activity appears considerably lower than expected on liking received. Both the second and the third men are higher on the average than he. Is the top man doing something to lose likes and provoke dislikes? Here one notes the dislike curve. The differences are small and probably not significant but they suggest that the top man is possibly the highest on dislikes received. Liking seems to be centering on the second and third man in activity, and they both seem to be lower than expected on idea ranking. Can it be that these men are tending to avoid too heavy an emphasis on the task area for fear of being disliked?

On further investigation of this problem it turned out that something happened in groups over a series of four sessions that was equally thought-provoking. In the first sessions, if a given man held top position on the idea ranking by his fellow members, the probability was about 50-50 that he would *also* hold a top position on a likeability ranking. But in the second meeting the probability dropped markedly, and by the fourth meeting was only about one in ten. The percentage of cases in which the same man held the top position on liking and idea rankings at the same time, divided by session, may be charted as follows:

Sessions

1	2	3	4
56.5	12.0	20.0	8.5

Could it be that there was something about arriving in a top-status position, owing to technical contribution to the task problems of the group, that tended to "lose friends and alienate people"? If so, was another man likely to arise who paid more attention to the social-emotional problems of the group and so tended to collect more liking? The idea that this happens with sufficient frequency that it can be viewed as typical may be called "the hypothesis of two complementary leaders."

The Hypothesis of Two Complementary Leaders

Why, if at all, should groups tend to have two complementary leaders, one a task specialist, the other a social-emotional specialist?[7] Perhaps it would be helpful to look at the interaction of men highest on idea ranking received but not highest on liking received, and vice versa. It may be that men of these two types behave differently and the differences in behavior may give us some clues as to the reasons for the differences.

Table 1 shows the composite profiles of 44 matched session-pairs[8] of idea men (who were not best liked in their group)

[7] A theory is advanced in Robert F. Bales and Philip E. Slater, "Role Differentiation in Small Decision-making Groups," Ch. V in Talcott Parsons *et al.* (eds.), *Family, Socialization, and Interaction Process* (Glencoe, Ill.: Free Press, 1955).

[8] Although the number of *sessions* was 44, the number of separate individuals involved was not 88, since each group ran over four sessions, and some individuals were in the same position more than once.

Table 1[a]

Composite Profiles in Percentages of 44 Top Men on Idea Ranking and 44 Top Men on Like Ranking for the Same Sessions.

	Interaction category	*Initiated* Idea men	*Initiated* Best-liked men	*Received* Idea men	*Received* Best-liked men
Area A: Positive reactions	Shows solidarity	3.68	4.41	2.57	3.15
	Shows tension release	5.15	6.98	7.95	9.20
	Shows agreement	14.42	16.83	23.29	18.27
Area B: Problem-solving attempts	Gives suggestion	8.97	6.81	7.01	7.22
	Gives opinion	32.74	28.69	25.52	31.09
	Gives orientation	18.54	17.91	14.06	14.54
Area C: Questions	Asks orientation	3.04	3.71	3.62	2.80
	Asks opinion	1.84	2.94	1.94	1.74
	Asks suggestion	.93	1.33	.85	.84
Area D: Negative reactions	Shows disagreement	8.04	7.60	10.65	9.35
	Shows tension increase	1.92	2.16	1.59	1.35
	Shows antagonism	.73	.63	.95	.45

[a]From Philip E. Slater, "Role Differentiation in Small Groups," *Am. Sociol. Rev.,* 1955, XX, 305.

and best-liked men (who were not top in idea ranking). Slater, from whose paper the table is taken, comments: "The most salient general difference in Table 1 is the tendency for the Idea man to initiate interaction more heavily in Area B (Problem-Solving Attempts) and the Best-liked man in Area A (Positive Reactions). The Idea man also seems to disagree somewhat more, and show a little more antagonism, while the Best-liked man asks more questions and shows more tension."[9]

On the receiving end, the situation is largely reversed, with the idea man receiving more agreement, questions, and negative reactions, while the best-liked man receives more problem-solving attempts, and more solidarity and tension release. The general picture is thus one of specialization and complementarity, with the idea man concentrating on the task and playing a more aggressive role, while the best-liked man concentrates more on social-emotional problems, giving rewards and playing a more passive role.

The kind of complementarity that shows in the behavior, then, is a kind that occurs in short interchanges in conversations where a problem-solving attempt by one person is followed by an agreement or disagreement from some other, or where a pleasant remark or a joke by one is followed by a smile or a laugh from the other. Such a division of

[9]Slater, *op. cit.* in footnote to Table 1. It is not possible to state that all of the detailed differences indicated are significant, because rates in the various categories are interdependent. However, Slater shows that the two types are in general significantly different from each other.

labor by type of act is very common and easily recognized. There may or may not be a specialization so that one person continues to produce more of one form of behavior than the other.

But now consider an important fact. Almost exactly the same sort of difference in interaction profile tends to be found between high participators and low participators,[10] even if one ignores the idea and like ratings. High participators tend to specialize in problem-solving attempts, low participators tend to specialize in positive or negative reactions or questions. Moreover, the proportion of problem-solving attempts increases when a man is placed with lower participators and decreases when he is working with higher participators.[11] What do these facts suggest?

For one thing, these facts seem to imply that the qualitative differences in the type of act attributed to a given person may be more or less forced by the tendency of others in the group to talk a little or a great deal, thus giving him an opportunity to make the problem-solving attempts or leaving him only in a position to respond to the quicker or more valuable proposals of others.

Insofar as the ratings a man receives are based on the way he behaves, the ratings others give him will surely be dependent on how much he talks. Let us suppose that a man can receive a high rating on ideas only if he makes many problem-solving attempts. He can do this only by talking a good deal. Then, to receive a high rating on ideas he will have to talk a lot. Or, conversely, let us suppose that a man can receive a high rating on liking only if he rewards others by positive reactions. He can do this only if he permits them to make many problem-solving attempts, which in turn requires that he let the other(s) talk a lot. Then, to receive a high rating on liking he will have to talk less.

This line of reasoning seems to fit with the facts so far presented and, moreover, has a certain plausibility in terms of common organizational arrangements. The husband and wife in many families seem to play conplementary roles of the sort described. Many administrators find cases from their experience where organizations in fact have two leaders, one who specializes on the task side, one on the social-emotional side. It is a kind of political maxim that it is almost impossible to elect the person who is technically best suited for an office—he is generally not popular enough. Surely there must be many persons in leadership positions who welcome any theory that explains to them that their lack of popularity is no fault of their own but a result of a specialization that is in the nature of things.

The problem now is that it might be inferred from this ideological version of the theory that there is no essential distinction between sheer activity and ratings received on goodness of ideas and, moreover, that there is a negative correlation between these two and liking received. Is it true that leaders must choose between task effectiveness and popularity?

[10]See Edgar F. Borgatta and Robert F. Bales, "Interaction of Individuals in Reconstituted Groups." *Sociometry*, 1953, XVI, 302–320.
[11]*Op. cit.*

The Hypothesis of Three Orthogonal Factors

Fortunately, a number of studies in the literature bear on this question and the results of a number of researchers tend to converge on an answer. When members of small groups are asked to rate and choose each other on a wide variety of descriptive criteria or are assessed by observers, three factors or distinct dimensions generally tend to appear.

Carter[12] indicates the frequency with which these factors are found in reviewing a series of factor analytic studies, such as those of Couch and himself, Sakoda, Wherry, and Clark.[13] A recent study by Wispe[14] may be added to the list.

Carter describes the factors as follows:

Factor 1. *Individual prominence and achievement:* behaviors of the individual related to his efforts to stand out from others and individually achieve various personal goals.

Factor 2. *Aiding attainment by the group:* behaviors of the individual related to his efforts to assist the group in achieving goals toward which the group is oriented.

Factor 3. *Sociability:* behaviors of the individual related to his efforts to establish and maintain cordial and socially satisfying relations with other group members.

These factors seem to represent underlying dimensions in the evaluations persons make of each other, whether as observers or as fellow group members. It may be that the best way of looking at these factors is not as personality traits but as frameworks in which the perceiver responds to personality traits of others.

But the important thing to note is that in these studies the three factors, which I shall call "activity," "task ability," and "likeability," are not, in general, mutually exclusive: a high standing on one does not preclude or interfere with a high standing on the other. Nor are they mutually supportive in general but, rather, they tend to be uncorrelated.

The fact that they are uncorrelated in general does not necessarily mean, of course, that there are no dynamic relationships between the phenomena represented by the factors. It means that there is no simple linear relationship that tends to be found over all populations, so that knowing where a man stands on one does not allow for a prediction of his standing on either or both of the others. If there are dynamic relationships between the factors they must be more

[12] Launor F. Carter, "Recording and Evaluating the Performance of Individuals as Members of Small Groups," *Personn. Psychol.*, 1954, **VII**, 477–484.

[13] Arthur S. Couch and Launor F. Carter, "A Factorial Study of the Rated Behavior of Group Members," Paper read at Eastern Psychological Association, March 1952; J. M. Sakoda, "Factor Analysis of OSS Situational Tests," *J. Abnorm. & Soc. Psychol.*, 1952, **XLVII**, 843–852; R. J. Wherry, *Factor Analysis of Officer Qualification Form QCL–2B* (Columbus: Ohio State University Research Foundation, 1950); R. A. Clark, "Analyzing the Group Structure of Combat Rifle Squads," *Am. Psychologist*, 1953, **VIII**, 333.

[14] Lauren G. Wispe, "A Sociometric Analysis of Conflicting Role-Expectations," *Am. J. Sociol.*, **LXI** (1955), 134–137.

complicated, nonlinear, or circumstantial. What suggestions of relationship are there left?

The Hypothesis of Individual Differences in Overtalking

Although it is not true that simply by talking a great deal does one guarantee a high rating on the quality of his ideas, it is still probably true that in groups of the sort we were studying it is very difficult to make a substantial contribution to the task without talking a great deal, especially in the first meeting, and overtalking may be resented by other members as a threat to their own status and a frustration of their own desire to talk. Results of other experimenters provided some findings that are congruent with this line of thought. Let us look for a moment at some of these results.

Leavitt and Mueller[15] explored the effect of one-way communication in a restricted communication situation where the receiver of the information is given no opportunity to "feed back" acknowledgements, questions, or negative reactions to the sender. They find that an initial reaction of hostility toward the sender tends to appear.

Thibaut and Coules[16] find that receivers who are not permitted to communicate to a person who has sent them an act of hostility show less post experimental friendliness to the sender than those permitted to reply.

A peripheral position in a restricted network approximates in some ways the position of a receiver with no opportunity for feedback. In an experiment where members were allowed to communicate only in written form through set channels on a task of assembling information, Leavitt[17] finds that members in peripheral positions are less well satisfied with their jobs than those in central positions.

These results suggested to us that the relatively low average of likeability preferences received by top participators might be due to the presence of some men in the total population of top men who overtalk, in the sense that they do not allow an appropriate amount of feedback of objections, qualifications, questions, and countersuggestions to occur. Our method of observation allowed us to examine the amount of interaction a given man received in relation to the amount he initiated. We thus arrived at the hypothesis that the ratio of interaction received to that initiated might help distinguish between those top interactors who were proportionately well liked and those who were not.

In general, as has been indicated, activity, task-ability ratings, and liking ratings appear in many studies as orthogonal factors, uncorrelated with each other over the total population assessed. It is important to recognize, however, that subparts of a population, or a

[15] H. J. Leavitt and R. A. H. Mueller, "Some Effects of Feedback on Communication," *Hum. Relat.*, 1951, **IV**, 401–410. See Research Example for Exercise 12.

[16] J. W. Thibaut and J. Coules, "The Role of Communication in the Reduction of Interpersonal Hostility," *J. Abnorm. & Soc. Psychol.*, 1952, **XLVII**, 770–777.

[17] H. J. Leavitt, "Some Effects of Certain Communication Patterns on Group Performance," *J. Abnorm. & Soc. Psychol.*, 1951, **46**, 38–50.

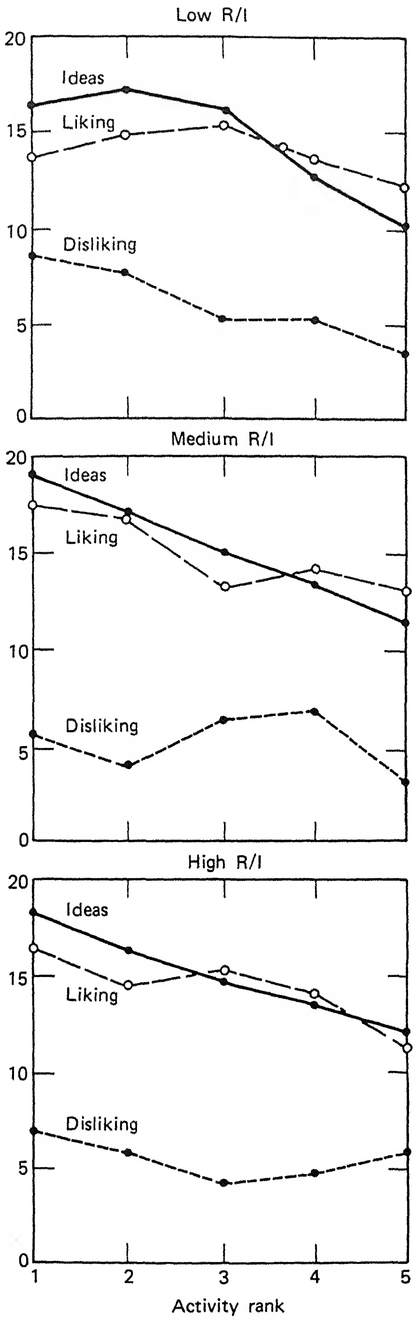


Figure 3. Average ratings* received on ideas, liking, and disliking by men of each activity rank, according to their feedback ratio (R/I). (Condensed from Robert F. Bales, "Task Status and Likeability as a Function of Talking and Listening in Decision-Making Groups," in Leonard D. White (Ed.). *The State of the Social Sciences.*

*Each entry at a given activity rank is the mean for 10 persons. The idea index is not actually a rating, but an index obtained by adding rankings received (including self-rankings) and subtracting the total from the highest possible, 25. The like and dislike indexes are average ratings, with the highest possible, 28.

different population, may show the variables related in a different way. It is the possibility that subparts of our population may show different relationships of these variables that we now explore.

We first make a basic division of the

population according to the rank of each person within his own group on the gross amount of participation he initiated and call this his activity. Five ranks are thus recognized, since the groups were five-man groups.

The second division of the population is made within each rank. All the men of each rank are divided into three subpopulations according to their own ratio of amount of participation received from others to the amount of participation they initiate. This ratio is known as the *R/I*, or the feedback ratio. Within each rank, then, there are three subpopulations of ten men each, low, medium, and high on the feedback ratio.

Figure 3 shows the average values of ratings or ranking received for each of the subpopulations of ten men on liking, disliking, and ideas. The ratings or rankings were given to each man by his four fellow group members and have been converted for plotting in such a way that high numbers mean high average rankings received.

The point of greatest interest is the difference in the relations of liking to activity when the feedback ratio is taken into account. Figure 3 indicates that among the third of the population with a low feedback ratio, the top two men seem definitely lower than would be expected if liking received increased linearly in proportion to activity. The correlation between activity and liking received is near zero.

However, both the medium *R/I* and the high *R/I* thirds show a positive correlation. From these data it is still plausible to suppose that the top man even in the high *R/I* third shows a little less liking received than one would expect. But the effect is slight.

The data obtained by asking about dislikes present essentially the same picture. The highest participators among the third of the population with the lowest feedback ratio not only are less well liked but are more disliked than their less active colleagues in the same subpopulation. In this third of the population, the more the person talks, the more he is disliked. But in the opposite third of the population, those who have a high feedback ratio, there is no relation between how much a man talks and how much he is disliked.

With regard to idea rankings received, there is a definite indication that the highest participators in the third of the population with the low feedback ratio tend to suffer on idea rankings received, as they do on liking received, although the effect is not so marked. This effect seems to disappear completely in the medium *R/I* and high *R/I* groups.

It is plain, however, that there is an appreciable linear correlation between activity and idea rankings received over the total of the three subpopulations. This finding thus differs from other studies which find these two variables to be generally orthogonal. We attribute the correlation in our groups at least partly to the fact that we are dealing in this study with data from first meetings entirely. Data on groups running over four sessions indicate that this correlation tends to fall over time, especially in groups where the initial consensus as to who has the best ideas is low.[18] The correlation between ideas and liking also tends to fall as indicated above in Table

[18] Philip E. Slater, "Role Differentiation in Small Groups," *Am. Sociol. Rev.*, 1955, XX, 300–310.

1. In short, the three factors tend to separate out as independent more clearly in later meetings than in the first.

To summarize briefly: In the groups in this total sample there is only a weak correlation between liking received and activity, providing one makes no breakdown into subpopulations. But for about one third of the population there is a positive and linear correlation between how much a man talks and how well he is liked. This is the third, who receive more interaction in proportion to the amount they initiate, that is, who have a high feedback ratio. The falling-off of liking received among the individuals who talk the most in total population is attributable especially to the other extreme third of the population, who talk proportionately most above the amount they receive. The same may be said for their rankings.

Conclusion

It appears that activity, task-ability ratings, and likeability ratings should be treated as three distinct factors, since over a large population of members, meetings, and groups they tend to be uncorrelated with each other. If one accepts this assumption a simple and very useful classification of role types in small groups suggests itself.

1. A member who is high on all three of the factors corresponds to the traditional conception of the good leader, or the "great man." Such men are found, but, if the factors are uncorrelated, are rare.

2. A member who is high on activity and task-ability ratings but less high on likeability is a familiar type who may be called the "task specialist." This type is not so rare as the first type and may operate effectively in cooperation with the third.

3. A member who is high on likeability but less high on activity and task ability may be called the "social specialist." This type is much less rare than the first type and groups which operate under the dual leadership of a man of this type and of the second type are common.

4. A member who is high on activity but relatively low on task ability and likeability ratings may be called an "overactive deviant." This type is not rare. This is the person who, in the leadership literature, is said to show "domination" rather than "leadership."

5. A member who is low on all three may be called an "underactive deviant" and may indeed be a kind of scapegoat. On the assumption that the factors are uncorrelated this type should be as rare as the first type, but since the lack of correlation traces mainly to discrepancies at the upper end of the scales, this type is not actually so rare as the first type and is, in fact, probably very common.

Logically, of course, one can distinguish many additional types. Those mentioned, however, have a certain intuitive distinctness and for present purposes serve to summarize and harmonize the various views on role differentiation that have been examined in this paper.

Exercise 2

Communication and Feedback

Much of the information transmitted in today's mass society reaches us via one-way broadcast systems. Television viewers are unable to signal an announcer when they do not understand him, and college students attending a large class have little opportunity to question the lecturer or to ask him to clarify a point. Because two-way communication in mass audience situations such as these is usually impossible or uneconomical, most lecturers and television writers prepare their messages very carefully. They know that they will not be able to adjust a message when it is in progress because they will not be able to receive feedback from the audience.

In many situations where it is structurally possible to build feedback into a communication system, the status or personality of the communicator prevents adequate feedback or questioning. For example, a listener may feel uncomfortable asking his host to repeat a joke, or an employee may fear that he will appear stupid if he asks the boss to repeat some instructions.

The absence of feedback can frustrate the sender as well as the receiver of a communication. Without feedback the sender does not know that his message got across, and the receiver does not know if he received the message exactly as it was intended. With adequate feedback channels (1) the receiver is assured that at the end of a message he will understand what the sender wanted to communicate and (2) the sender knows that his message has been accurately received. For messages where accuracy is essential, feedback loops as diagrammed in Fig. 1 are of utmost importance.

In Fig. 1, the messages are numbered sequentially from top to bottom. The receiver continues to check the message until he receives a confirmation that he has correctly understood the intended message. After the sender

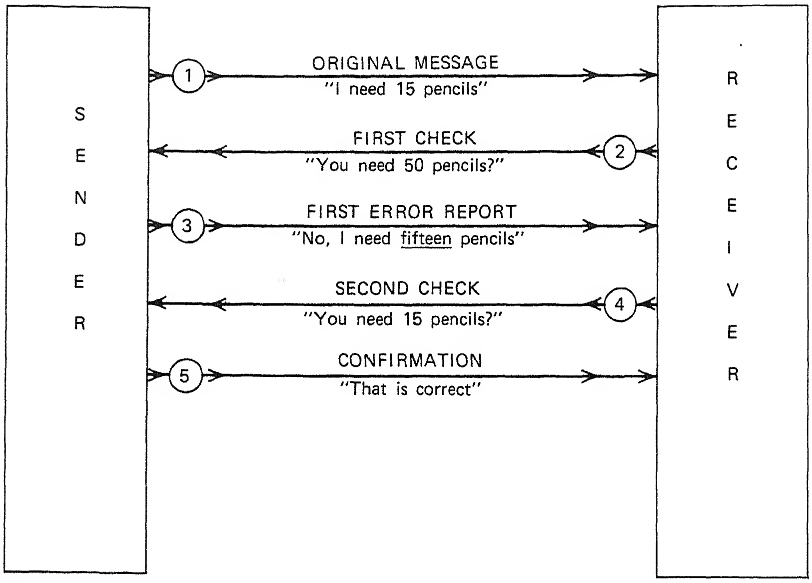


Figure 1. Feedback loops.

transmits a confirmation he may terminate or he may send another message. The minimum number of transmissions to assure understanding is three—the original message, a check, and a confirmation. Each error requires an additional feedback loop of two messages—a check answered by either a confirmation or an error report.

This exercise, which is adapted from Leavitt and Mueller (1951 reprinted as a Research Example p. 347), compares three kinds of communication situations—one-way or zero feedback, two-way or free feedback, and a limited yes-no feedback condition—on three dependent measures: accuracy, speed, and confidence.

Procedure

First find a friend to act as the subject or communication receiver. You will need several blank sheets of standard 8½" X 11" paper and a watch with a second hand. Your job as communicator will be to describe verbally the three patterns of rectangles that are shown on page 341 as accurately as possible. Give the subject a piece of paper and instruct him to draw the pattern as you describe it. Arrange yourselves so that he cannot see the pattern and so that you cannot see his drawing. Do not use gestures, drawings, or objects—words only. Keep time for each trial and decide that the trial is completed when you feel you have communicated the pattern to the subject as best you can.

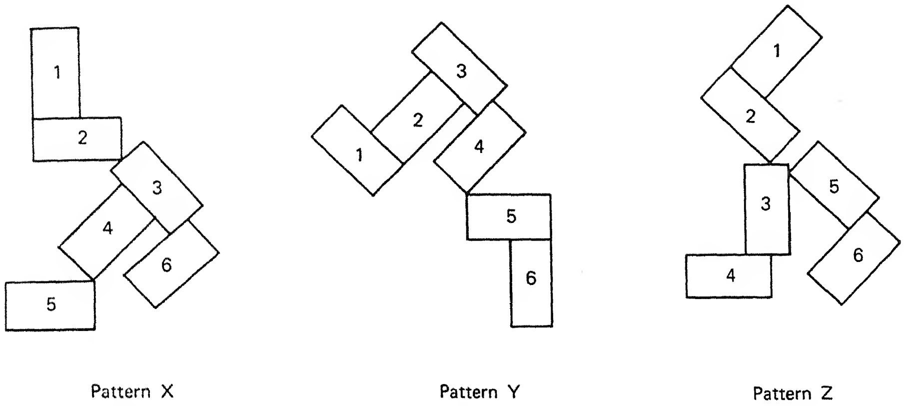


Figure 2. Rectangle patterns (to be communicated to subject).

VERBAL FEEDBACK CONDITIONS:

A. Zero Feedback

The subject is not permitted to ask questions or to make comments (do not allow even grunts or sighs).

B. Yes-No Feedback

The subject is permitted to answer any questions you ask with "yes" or "no." You are allowed to initiate communication whenever you wish by asking a question; the subject is restricted to responding with a "yes" or "no." Subjects may not ask questions or make comments.

C. Free Feedback

The subject is permitted to interrupt and ask questions whenever he wishes. You may ask questions whenever you feel they are necessary or useful.

VISUAL FEEDBACK CONDITIONS:

D. No Visual Contact

Do not permit the subject to see you during the communication period. You may use a screen or partition, speak from the next room, or communicate with the subject over a telephone or intercom.

E. Face-to-Face

You should sit facing the subject at a distance of from three to eight feet. Do not let the subject see the patterns to be communicated.

The three levels of the verbal feedback variable, A, B, and C, and the two levels of the visual feedback variable, D and E, are combined in a 2 X 3 factorial experimental design yielding six different conditions as illustrated in Table 1.

Table 1
Experimental Design

		Verbal Feedback		
		A ZERO	B YES-NO	C FREE
Visual Feedback	D No Visual Contact			
	E Face-to-Face			

Always start with Pattern *X* and communicate it to your subject; then proceed to Patterns *Y* and *Z*. Each class member will test only one subject and this subject will attempt to reproduce the three patterns, always in *X, Y, Z* order. Each subject will be tested under all three verbal feedback conditions; however, there will be several different orders. If the verbal feedback conditions were always presented in the same A, B, C order, differences among the conditions on one or more of the measures could be attributed to the order of presentation of the verbal feedback treatment. For example, if the zero feedback condition were always presented last, it could yield consistently high scores because of a practice effect or low scores because of fatigue or boredom. To eliminate this confounding (see Introduction, page 17ff. for a discussion of confounding) of order and feedback condition as an explanation for any differences, presentation order can be counterbalanced so that the three verbal feedback conditions will be run in all possible six orders: ABC, ACB, BAC, BCA, CAB, and CBA. If only a few experimenters are available, the group may be divided as follows: if your *last* name begins with A through K start with condition A (zero feedback); then do B and C. If your *last* name begins with L through Z, start with condition C (free

feedback), then do B and A. The design will not be perfectly counterbalanced with only two orders; however, the results may suggest that order makes a difference. In this experiment, verbal feedback is termed a "within-subject" variable because each subject is exposed to all (three) of the verbal feedback conditions.

On the other hand, visual feedback is labeled a "between-subjects" variable because each subject will participate in only one of the (two) visual conditions. If your *first* name begins with A through K, run your subject under face-to-face conditions. If your *first* name begins with L through Z, do not allow your subject to see you during the communication periods.

Record the following data on Data Sheet 1 immediately after each trial.

1. Accuracy: The first rectangle is scored as correct if it is properly oriented on the page. Other rectangles are placed correctly if they bear the correct relationship to the preceding rectangle. Each correct rectangle counts for one point; thus, the accuracy score can range from 0 to 6.

2. Speed: Record the total communication time in seconds including any pauses and questions from the subject.

3. Confidence: After each trial, and before looking at the subject's drawing, rate your own confidence that the pattern has been perfectly reproduced. 5 indicates complete confidence; 4—somewhat confident; 3—slightly confident; 2—unsure; 1 indicates certainty that the pattern was *not* drawn correctly. Then, read this rating scale to the subject and ask him to rate his confidence that he has drawn the pattern perfectly.

Discussion

If, as suggested, students tested only one subject, the class will need to collate the data before proceeding. Accuracy, time, and confidence scores can be segregated into the 6 feedback conditions as illustrated in Table 1. Means calculated across the two rows in Table 1 can be compared by using a t-test (Appendix A) to determine whether the visual feedback condition makes a difference. The means for the three columns representing verbal feedback conditions however, cannot be compared by using the t-test described in Appendix A. That t-test requires that the observations made under each treatment condition be independent or uncorrelated with one another. In this exercise, the same person was exposed to each of the three verbal feedback conditions; therefore, the observations in the different verbal feedback treatments are not independent of each other. It may be sufficient to simply inspect the average scores of subjects within each of these three groups, or you may wish to use a t-test for correlated means, which may be found in most elementary statistic books.

Those students who have studied statistics may recognize that analysis of variance should be employed here to calculate effects of both variables and the interaction or relationship between them; however, the t-tests will suffice for purposes of this exercise. As mentioned in the procedure section, differences between the columns could also be because of practice or fatigue effects if the design was not perfectly counterbalanced for order. Note that the design is *not* counterbalanced for pattern; pattern X was always presented first followed by Y and Z. What confounding effects might the pattern have? If the pattern difficulty could cause differences that confound the data, how can this be remedied in a more complicated experimental design either by counterbalancing or by using a between-subjects design? Whether or not you have counterbalanced, you might want to find out if pattern and order have effects by segregating your data accordingly.

If you have determined that the data of theoretical interest (differences among feedback conditions) are not likely to be confounded or open to alternative explanations, discuss the results you have obtained. What implications, if any, do these data have for building communication systems, for organizing meetings, for designing classrooms? You may find it helpful to read Leavitt and Meuller's discussion of their findings in the Research Example, p. 347.

Data Sheet 1

Subject ____________________

Trial	Pattern	Verbal Feedback Condition	Accuracy Score (0–6)	Time (Secs.)	Experimenter Confidence Rating (1=complete confidence)	Subject Confidence Rating (1=complete confidence)
1.	X					
2.	Y					
3.	Z					
Average for 3 trials						

Circle one:

D. NO VISUAL CONTACT

E. FACE-TO-FACE

Comments:

Research Example

Some Effects of Feedback on Communication[1]

HAROLD J. LEAVITT AND
RONALD A. H. MUELLER

Introduction

The experiments reported here are concerned with the transmission of information from person *A* to person or persons *B*. Our problem deals with only one of the many relevant variables, the variable of feedback. The question becomes: how is the transmission of information from *A* to *B* influenced by the return of information from *B* to *A*? It is apparently taken for granted in industry, in the lecture hall, and in radio that it is both possible and efficient to transmit information from *A* to *B* without simultaneous feedback from *B* to *A*. On the other hand, the information theories of the cyberneticists and, to some extent, trial and error concepts in learning theory suggest that for *A* to hit successfully some target, *B*, requires that *A* be constantly informed of *A*'s own progress. The servomechanism needs a sensory system that is capable of transmitting cues about the errors of its own motor system. The human being learning some motor skill apparently utilizes the same process. But when the human being (*A*) seeks to transmit information to another human being (*B*), *A*'s own sensory system is hardly an adequate source of information *unless* *B* takes some action which will help *A* to keep informed of *A*'s own progress. If *A* were trying to hit *B* with a brick, *A*'s eyes combined with an inactive *B* would probably be adequate to permit *A* to hit his target after several trials. But if *A* seeks to hit *B* with information, he will probably be more successful if *B* helps to provide some cues which *A*'s own sensory system cannot pick up directly. In other words, where communication between *A* and *B* is the goal, feedback, in the form of verbal or expressive language, should make for greater effectiveness.

If we take the human memory mechanism into account, we need not require that there be *contemporaneous* feedback between *A* and *B*. It may not even be

SOURCE. *Human Relations,* 1951, 4, 401–410.

[1] Readers familiar with the work of Professor Alex Bavelas and his group at M.I.T. will doubtless correctly recognize that many of the theoretical and experimental ideas in this research had their origins in that group. We are most grateful to Dr. Bavelas for both his direct and indirect help.

necessary that there be any feedback from B_2 if feedback from a similar B_1 has already occurred. The practice sessions of the past may have provided enough feedback to permit one to hit his present target accurately. Language, for example, may be thought of as a tool originally learned with feedback, but currently useful in a multitude of situations without simultaneous feedback to help us at least to get within range of our targets. But if the material to be communicated is relatively new and relatively precise, previously learned language may not be enough. Accurate transmission may require some additional contemporaneous feedback.

In addition to this hypothesis that contemporaneous feedback should increase the accuracy of transmission of information from *A* to *B*, is the hypothesis that the completion of the *AB* circuit produces other effects on the *AB* relationship. Feedback from both *A* and *B* can increase the certainty of *B* that he is getting the intended information, and the certainty of *A* that he is getting it across. This increase in certainty, assuming motivated participants, should have some effect on feelings of frustration or achievement and, hence, on the feelings of hostility or security that pervade the relationship.

Our purpose, then, in these experiments is to try to test these hypotheses; to try to determine experimentally the effects of feedback (or the absence of feedback) on certain kinds of *A* to *B* communications.

Experiment I

WHAT ARE THE EFFECTS OF PROGRESSIVE LEVELS OF FEEDBACK?

We chose as our material-to-be-communicated in these experiments a series of geometric patterns. The patterns were all composed of six equal rectangular elements, but the relationships of the elements to one another differed from pattern to pattern (see *Fig. 1A* for sample pattern). *A*'s (the instructor's) job was to describe orally one of

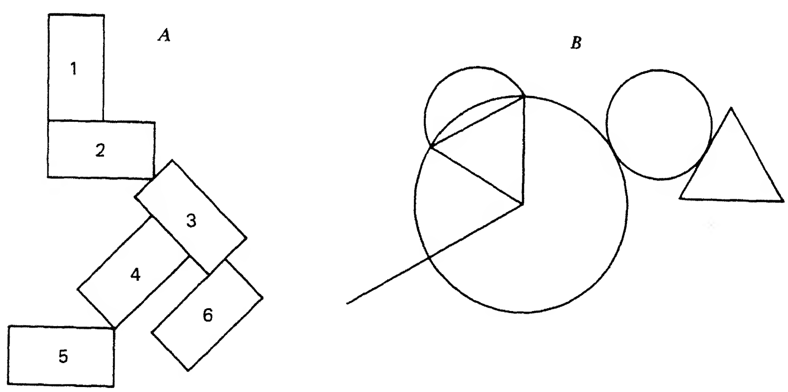


Figure 1. Sample problems, (*A*) Sample of problems used in Experiment I. (*B*) Sample of problems used in Experiment II.

Table 1

Design of Experiment I

Pattern No.	*1*	*2*	*3*	*4*	*5*	*6*	*7*	*8*
Class 1:	zero	V–A	Y–N	free	zero	V–A	Y–N	free
		(Instructor X)				(Instructor Y)		
Class 2:	V–A	Y–N	free	zero	V–A	Y–N	free	zero
		(Instructor Y)				(Instructor X)		
Class 3:	Y–N	free	zero	V–A	Y–N	free	zero	V–A
		(Instructor X)				(Instructor Y)		
Class 4:	free	zero	V–A	Y–N	free	zero	V–A	Y–N
		(Instructor Y)				(Instructor X)		

these abstract patterns to the members of his class as accurately as possible, accuracy to be measured from the students' reproductions of the described (but unseen) patterns.

Two instructors were used, and four groups of students (total student N = 80), with each instructor describing four patterns to each student group. There were four conditions of feedback: 1. *Zero feedback* in which instructors sat behind a movable blackboard to describe the patterns. No questions or noises were permitted from the students. 2. The *visible audience* condition in which students and instructor could see one another but no speaking by students was allowed. 3. A *yes-no* condition in which the visible audience was permitted to say only yes or no in reponse to questions from the instructor. And 4. a *free feedback* situation in which students were permitted to ask questions, interrupt, etc.

With the use of a kind of Latin Square arrangement it was possible then to have each instructor use each condition of feedback in a different order. (See *Table 1.*)

Besides reproducing the test patterns, students were asked to estimate their confidence in the correctness of their answers and, after the last pattern, to indicate the feedback condition they found most comfortable. We also timed the description of each pattern.

All students were given the same instructions at the beginning of the class period. They were told that the experiment was a test of their ability to understand instructions, and that they were to work as rapidly and as accurately as possible. Both instructors had had some previous experience in describing similar patterns, and both had participated in the construction of the test patterns.

Students' papers were scored for accuracy on a scale from 0 to 6. A particular rectangular element was scored correct if it bore the correct relationship to the preceding element. The first element was scored correct if it was correctly oriented on the page.

RESULTS

1. Accuracy

The mean accuracy score for *all* patterns increased steadily from *zero* to *free*

feedback.. With *zero feedback* the mean was 4.7 out of a possible 6. The range of means for the eight different patterns given under this condition was 3.1 to 5.9. Under the *visible audience* condition the mean score was 5.3 with a range from 4.5 to 5.9. Under the *yes-no* condition the mean score was 5.5, the range 5.0 to 5.8. With *free feedback* the mean was 5.6 and the range 5.1 to 6.0

2. Confidence Level

Students' estimates of their own accuracy correlated closely with actual accuracy. For all patterns the mean confidence levels were: *zero feedback*, 4.6; *visible audience,* 5.3; *yes-no,* 5.6; *free feedback,* 5.5. No effects of experience could be detected. There was a tendency to favor one instructor for the *free feedback* situation and the other for all others. These differences were slight and may indicate a differential skill on the part of the instructors in handling the different feedback conditions.

3. Time

The mean time required to give instructions under the four conditions were: *zero feedback*, 229 seconds; *visible audience*, 249 seconds; *yes-no,* 316 seconds; *free feedback*, 363 seconds. Any decrease in time with experience is once again obscured by differences in difficulty. No clear-cut differences between instructors were apparent.

4. Other Observations

Both instructors noticed some rather interesting behavior under certain conditions. When using *free feedback*, both found that on some occasions the students utilized their opportunities to speak by speaking aggressively and with hostility. There were comments like: "That's impossible"; "Are you purposely trying to foul us up?"; "You said left, it has to be right"; and so on. These comments even flowed on to students' papers, when they wrote beside their patterns such comments as: "The teacher made mistakes on this one, I didn't." These hostile reactions seemed to occur only when the *free feedback* condition *followed* other conditions. Both instructors noticed too that their *free feedback* experience stood them in good stead in the *zero feedback* situations. A student in the *free feedback* situation might say, "Oh, it looks like an L." In the next use of that pattern the instructors would find themselves saying, "It looks like an L."

COMMENTARY

Although these data indicate that *free feedback* does yield more accurate results than the other conditions, some new questions arise. Can it not be argued that the *free feedback* method is more effective simply because it requires more time? Would the time required decrease if *free feedback* were used continuously? Does the *free feedback* method always put the teacher on the spot? Will he be attacked for his errors or lack of knowledge? Though free feedback may be helpful at first, is it of any use after the student and the teacher have had an opportunity to straighten out their language difficulties? Can the teacher improve just as much after a series of experiences without feedback as after a

series with feedback? Can we show continuous improvement in the course of several trials without feedback?

Experiment II

FEEDBACK VS. NO FEEDBACK

In an attempt to answer some of these questions we designed another series of experiments that seemed to permit the most efficient use of our limited supply of instructors and students. The purpose of these experiments was to compare the two extreme conditions, *free feedback* and *zero feedback*, over a longer series of trials.

METHOD

Using eight new geometric patterns, all made up of six elements (see *Fig. 1B*, p. 348), we selected ten instructors and ten separate groups of students, the groups ranging in size from six to twenty-four. Five of the instructors were members of the English Department of the Institute, one taught German, one economics, and three psychology. Four of the classes were speech classes, six were general psychology. For *three* pairs of instructors the procedure was as follows:

Instructor *A* faced class *A* with four patterns in sequence and *zero feedback.* Then instructor *B* faced class *A* with four new patterns in sequence and *free feedback*. Instructor *A* then faced class *B* with his original four patterns and *free feedback.* Then instructor *B* faced class *B* with his original four patterns and *zero feedback.* For the other two pairs of instructors the procedure was reversed, instructor *A* beginning with *free feedback.*

We again asked for confidence levels, from both the student and instructors.

RESULTS

1. Overall

The results of this experiment bear out the trend of the first. The mean student accuracy score for all *zero feedback* trials was 5.2 of a possible 6; the mean with *feedback* was 5.9. These means represent the students of ten instructors. The ranges for individual instructors were, with *zero feedback*, 3.8 to 5.8; with *free feedback*, 5.6 to 6.0. This difference between these means is significant as the 1% level.[2]

In students' confidence in their results, the data again correlate closely with accuracy. The mean for *zero feedback* is 5.0 with a range from 3.5 to 5.7, while for *free feedback* the mean is 5.8 and the range 5.4 to 6.0. These differences are also significant.

In terms of time required to describe each pattern, *free feedback* remains a more time-consuming process. The average time for *zero feedback* is 166 seconds with a range from 60 to 273. For *free feedback* the average time is 284 seconds with a range of 193 to 423. These differences too are significant.

Finally in our measure of teacher confidence, means were 4.5 with *zero feedback* and 5.0 with *free feedback* with respective ranges of 2.5 to 5.5 and 4.5 to 5.8. In all cases instructors were *less* confident than their students.

[2]Notations concerning significance levels are explained in the Introduction, pp. 15-16.

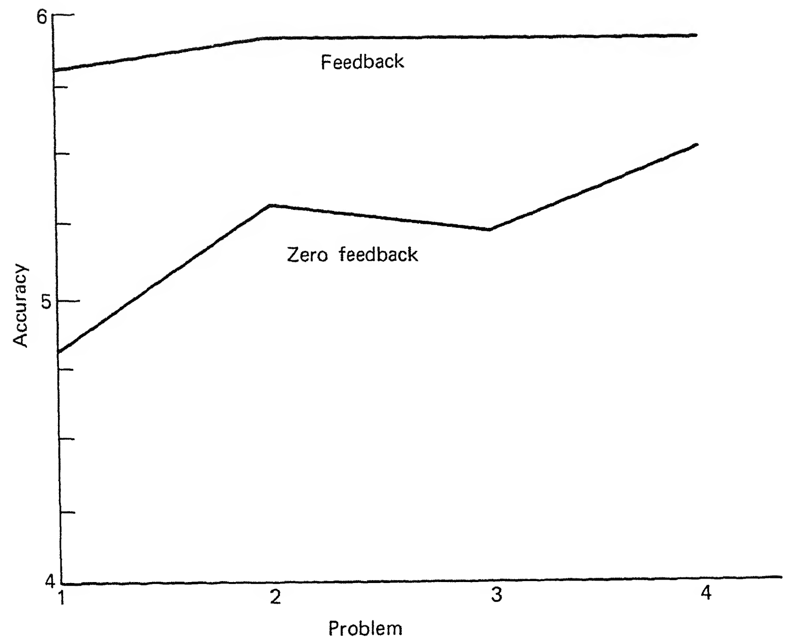


Figure 2. Accuracy. Each point represents the mean of 10 groups.

In every case individual instructors changed in the same direction as the means. Every instructor got better results with feedback than without, and every instructor took longer with feedback than without.

2. Effects of Experience

In *Figure 2* are shown curves representing the changes in accuracy from pattern to pattern. Each instructor, you will recall, described four patterns in sequence under conditions of *zero feedback* and then *free feedback.*

From these accuracy curves one can see that *free feedback* starts at almost the maximum level and stays there. *Zero feedback* changes in the direction of greater accuracy from trial to trial.

As far as time (*Fig. 3*) is concerned, the reverse is true. *Zero feedback* remains more or less constant, while *free feedback* time *declines progressively.*

There is at least one other way of analyzing the data that provides some rather interesting results. Our experimental design supplied us with data for all combinations of (a) inexperienced (with these patterns) and experienced instructors, and (b) inexperienced and experienced classes, working (c) with and without feedback. The data broken down this way indicate that instructors' experience is the most significant factor present. Differences between experienced and inexperienced instructors are always greater than between experienced and inexperienced classes. This difference holds for *zero feedback* only, since

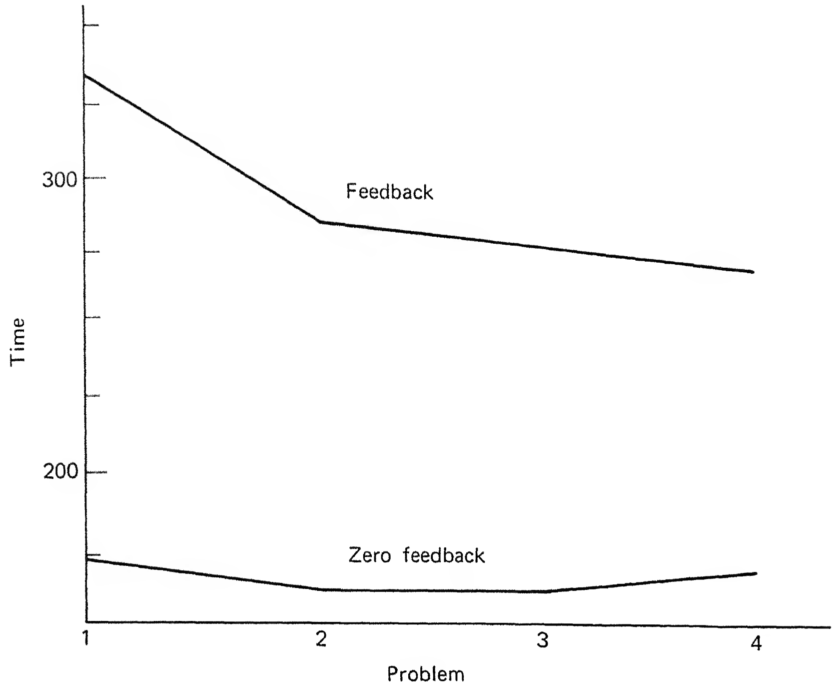


Figure 3. Time. Each point represents the mean of 10 groups.

with *free feedback* there are no perceptible differences among any of the different conditions.

3. Other Observations

One of our hypotheses in these experiments centered on the effects of feedback on the relationship between sender and receiver. We have no quantitative data that are relevant to this hypothesis, but we do have some observations that were astonishing in their consistency. These observations amounted to this. When an instructor faced a new class with *free feedback*, he got fairly rational feedback. That is, the students asked questions or asked for elaboration or repetition of a point. But when an instructor faced a class that had just been exposed to a *zero feedback* session, the instructor got an attack. The students asked lots of questions, but with barbs and implications about the instructor's (in)ability to do his job. The new instructor had innocently opened Pandora's box. This hostility did not last through more than one or two patterns, nor did it prevent the majority of students from expressing a preference for the *free feedback* method.

COMMENTARY

In a sense these experiments demonstrate the obvious. When a receiver *B* is free to ask questions he can get a better understanding of what the sender *A* is

trying to communicate. Moreover, with *free feedback* both the sender and the receiver can feel, correctly, more confident that they have respectively sent and received accurately. *Free feedback* requires more time, but there is some evidence that this time differential decreases with increased understanding between the sender and the receiver. Apparently the use of continuing *free feedback* could lead directly back into *zero feedback*, for once the common areas of misunderstanding have been clarified, contemporaneous feedback will no longer be necessary.

Apparently it is possible to improve communication skill with minimal feedback. The fourth *zero feedback* pattern is almost always more accurately sent than the first. This improvement can perhaps be thought of as a kind of personal feedback in which the instructor's own words are utilized to help him to increase his own effectiveness in the future. Much of it is no doubt empathetic, the instructor imagining himself in the receiver's place and correcting his sending as a consequence. Some of the improvement, however, may come from feedback which our experimental barriers failed to block out; feedback in the form of noises, sighs, shuffling of chairs. We do not know from these experiments whether or not an instructor using *zero feedback* could eventually reach the *free feedback* level of accuracy and confidence, but it is clear that under our experimental conditions he can improve over his own original *zero feedback* level.

Besides the findings about the direct effects of feedback, the data raise some questions about indirect effects. We observed in both experiments that *free feedback* after *zero feedback* is accompanied by hostility. This hostility was apparently an effect of the *zero feedback* experience. It lasts only through one or two *free feedback* trials. Why should this be so? We believe that the mechanism centers around the notion of "certainty." In our attempts to satisfy our needs we must be as certain as possible that we are successful. Uncertainty is frustrating. Without feedback uncertainty is prevalent.

In the same vein we noted that instructors' confidence is lower than students' confidence. We suggest that the instructor can be satisfied only by knowing that the receiver is getting the proper information. But the receiver can be satisfied by comparing his own work with the sender's directions. The receiver then has more information available against which to check his own progress toward his goal. Hence he can be more certain of his progress. But the sender is not sure of what the receiver is receiving. He can get *some* information with feedback, but almost none but his own empathy without feedback. Hence his certainty and confidence are low. These differential feelings of certainty, adequacy, and hostility seem to us to be the most significant differentials between our *free feedback* and *zero feedback* systems.

Summary and Conclusions

Since the scope of this research has been limited by the utilization of one kind of problem, one kind of sender-receiver situation, and a relatively short series of experiences, our conclusions must be severely circumscribed.

To summarize, we found that, within narrow limits: 1. a completion of the circuit between sender and receiver (feedback) increases the accuracy with which information is transmitted. 2. Feedback also increases receiver and sender confidence in what they have accomplished. 3. The cost of feedback is time. But the difference in time between *free feedback* and *zero feedback* appears to decrease. 4. A sender and a receiver can improve without what we have defined as feedback experience. 5. *Free feedback* experience improves subsequent *zero feedback* trials measurably. 6. Sender experience contributes more than receiver experience to improved accuracy of communication. 7. *Zero feedback* engenders some hostility in the receiver that becomes clearly perceptible when the situation *changes* from *zero* to *free feedback.* This hostility is short-lived, lasting through only one or two *free feedback* trials. 8. *Zero feedback* engenders doubt in the sender.

These findings support the hypothesis that *free feedback* is an aid to accuracy in interpersonal communication. *Free feedback* seems to permit the participants to learn a mutual language, which language once learned may obviate the necessity for further feedback.

The findings also support the hypothesis that the presence or absence of feedback affects the sender-receiver relationship. *Zero feedback* is accompanied by low confidence and hostility; *free feedback* is accompanied by high confidence and amity.

Exercise 3

Person Perception

As every trial lawyer recognizes, two witnesses seldom offer identical testimony about the same event. But why should two accounts of the same event differ so frequently? We usually assume that one or both witnesses have failed to make an accurate report of his perception. It may be that a witness unwittingly invents or eliminates information in order to fit with or assimilate to his own biases. Or he may perceive accurately, but quite deliberately choose to distort what he saw. Or perhaps different witnesses simply perceive a complex event from various vantage points and, hence, emphasize different aspects of the stimulus as was the case in the familiar story of the blind men, each of whom offered a different description of an elephant depending on whether they touched its trunk, tail, ear, or foot.

The greatest possibilities for disagreement between observers exist in the process of *person* perception, where the stimulus is another person. A person who is being judged, unlike an inanimate object in a physical perception situation, may choose to show different aspects of himself to various observers. Jim Foster may be one thing to his parents and quite a different stimulus object to his girl friend, his roommate, or his football teammates.

The fact that in this exercise raters will be required to make judgments about themselves as well as about other people, will create problems. Since most of us are careful about the kinds of things we say, particularly about ourselves, self-judgments may be even more susceptible to bias and distortion than judgments about other people; moreover, the Freudian concept of "defense mechanism" suggests that many people cannot describe themselves accurately even if they sincerely try. The basic data for this exercise are social judgments of yourself and another person. These judgments are to be

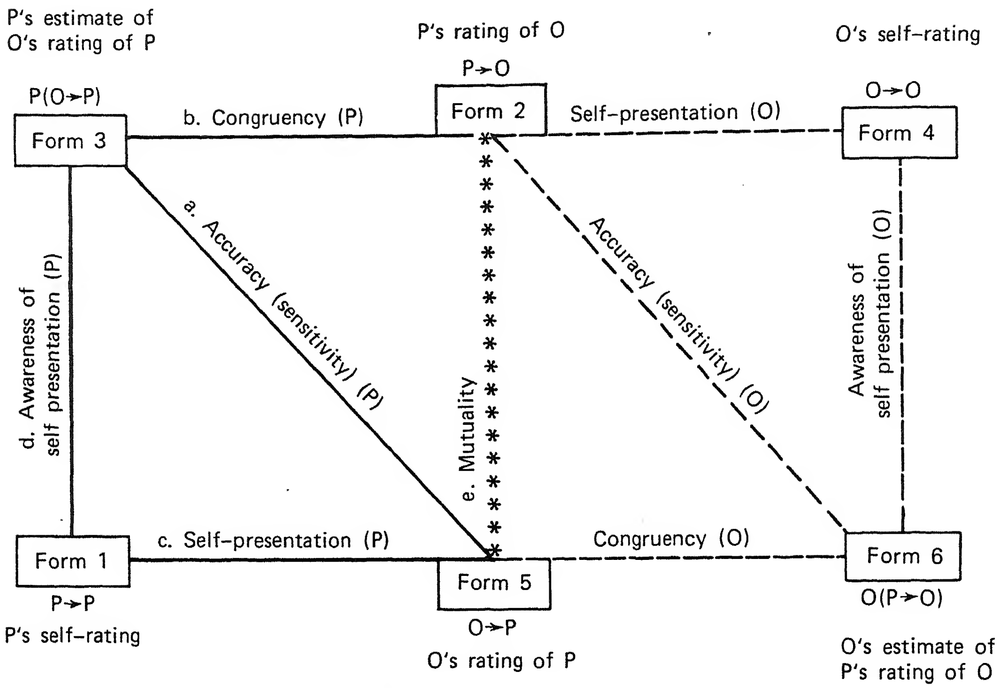


Figure 1. Diagram of social perception comparisons for two perceivers (P and O).

P = first person (you)
O = other person
______ = relevant comparison for P
\- - - - - = relevant comparison for O
******= mutual comparison

entered on six separate rating forms. You will fill out three of these forms and you will choose a partner who will fill out the remaining three forms. Both of you will rate (a) yourselves, (b) the other person, and (c) yourselves as you think the other person will describe you.

These six judgments yield nine comparisons which are diagrammed in Fig. 1. Each comparison will be explained in more detail in the Data Analysis section, but before reading further choose a friend, classmate, or even an enemy to serve as your partner. In order to conform with the notation commonly used in discussing person perception (Heider, 1959), you will be referred to as "P" (person) and your partner will be designated as "O" (others).

Procedure

Six Person Rating Forms are provided. Rate yourself on Form 1, rate the person you have chosen on Form 2, and on Form 3 describe yourself as you

think the other person will describe you. Then ask the other person to fill out Forms 4, 5, and 6. He should rate himself on Form 4, rate you on Form 5, and finally on Form 6, he should describe himself as he thinks you will describe him. *All forms should be completed before reading the Data Analysis section.* The other person's assistance will not be required for the remainder of this exercise, although you may wish to discuss the data analysis with him.

This exercise is somewhat unusual in that you will be collecting data about yourself. If you wish to keep your data confidential from the instructor, you may retain the Person Rating Forms. In this case, turn in only the Data Sheets on pages 368-370. It also might seem embarrassing to ask another person (0) to rate you, especially since 0 knows that you will see his ratings. You may be able to relieve some of this uneasiness by allowing O to see your ratings of him. Because studying yourself can create tension, you may wish to substitute someone other than yourself for P. If you decide to collect data from two other people, guarantee their anonymity by using code letters or initials on the Person Rating Forms. If you promise another person to keep his data anonymous that commitment is not to be taken lightly; be sure your procedures do, in fact, guarantee privacy.

Data Analysis

General Positive-Negative Ratings. Notice that the first four scale items A to D on the Person Rating Form are positive-negative in form. The polar adjectives at each end of the scales represent "good" and "bad" value judgments. On each scale one of the adjectives is clearly positive (Fair, Kind, Honest, Pleasant) and one is clearly negative (Unfair, Cruel, Dishonest, Unpleasant). These scales can be used to compare the relative amount of favorableness expressed in the three different types of rating forms, and these will be the first data analyzed.

Score items A to D on all the Person Rating Forms as follows: a mark on the blank closest to the more positive adjective (Fair, Kind, Honest, Pleasant) counts as 6; a mark on the blank closest to the most negative adjective (Unfair, Cruel, Dishonest, Unpleasant) counts as 0; the middle or neutral blank counts as 3; and so on. Score each of the four pairs of polar adjectives in this manner. Use Data Sheet 1 p. 367 to compute a general positive-negative rating for each of the six forms. The four numbers from scales A, B, C, and D from Person Rating Form 1 go in the left-hand column at the top of Data Sheet 1. Data from the other five forms should be entered in the appropriate columns on Data Sheet 1. Sum the four items for each scale by adding down columns to get the total evaluation score for each of the six forms. Divide the totals by 4 to obtain mean ratings.

Which of the three ratings of yourself (Form 1, 3, or 5) is most positive?

Most negative? If the ratings are quite different from each other, attempt to explain these differences. For example, your self-rating could be more negative than the other's (O's) rating of you for several reasons. Perhaps you have access to negative information about yourself that has been concealed from the other. Perhaps you (P) were simply being modest in appraising your own good qualities. Or perhaps you generally tend to rate everything (cars, people, omelets) more negatively than the other (O). This latter possibility is more tenable if the total positivity score for all of your P forms (1, 2, and 3) is lower than the same score for all of O's forms (4, 5, and 6). Low ratings can be attributed to either (1) a rater who always rates low, or (2) qualities of the persons being rated. See the optional further discussion on components of rating variance.

Comparisons Between Forms

Use the second set of four items from the Person Rating Forms (E, F, G, and H) to make the comparisons indicated in Fig. 1. The polar adjectives used in these items were not intended to be either positive or negative so that raters would be less susceptible to "social desirability" distortion, that is, on these items raters are less likely to make themselves or their partners appear favorable or unfavorable.

To facilitate tabulation, arbitrarily assign a score of 0 to a check mark placed in the extreme left-hand blank for items E, F, G, and H. A mark in the middle blank counts as 3, the extreme right-hand blank counts as 6, and so on. Transcribe the score for each item to the table at the top of Data Sheet 2. Notice that Data Sheet 2 (page 369) and the following discussion describe analyses for P (you) only. Analyses for O (the other) are symmetrical with those for P and can be calculated easily from Forms 2, 4, 5, and 6 if desired.

1. Accuracy (Forms 3 and 5). To determine how accurately you estimated the other person's ratings of you, compare your guesses about his ratings (Form 3) with his actual ratings (Form 5).

Look at your rating forms or at the table at the top of Data Sheet 2 and record the *discrepancies* in ratings for each pair of adjectives. For example, if you rated yourself as quite talkative (second blank from the right = 5), but the other person rated you as extremely talkative (extreme right-hand blank = 6), score a discrepancy of +1; and enter the score in the lower half of Data Sheet 2. The discrepancy would also be +1 if you rated yourself as quite talkative (5) and the other person rated you as only slightly talkative (third blank from the right = 4). All such discrepancy scores should be positive (+); none should have a negative sign preceding it because they represent the *absolute* difference between you and the other. Add across the

four discrepancies in row a, Data Sheet 2 (one for each item, E, F, G, and H) to obtain an accuracy or sensitivity score. A score of 0 indicates perfect accuracy; both you and the other person gave you exactly the same ratings. Were your guesses about the way the other person feels about you correct? If not, why were you inaccurate in estimating his feelings about you?

2. Congruency (Forms 2 and 3). In this comparison you will determine how closely your guesses about a person's feelings toward you are related to your feelings about him. Calculate the discrepancies between Forms 2 and 3 for each adjective pair and record them in row b, Data Sheet 2. The remaining three comparisons also should be recorded in the same manner (1 versus 5; 1 versus 3; and 2 versus 5). Congruency is almost always higher than Accuracy (Taguiri, Bruner, and Blake, 1958), which may indicate either that P tends to reciprocate O's perceived feelings, or that P's feeling toward O influences his guess about how O feels toward him. All of the links in Fig. 1 represent relationships between two ratings. However, no direction of causality is indicated, that is, these data do not tell us whether one of the ratings caused the other rating to shift. (See the discussion of causality in the Introduction, pages 11-14.)

3. Self-Presentation (Forms 1 and 5). Did O rate you as you feel you are? You should record the data and calculate discrepancies in Data Sheet 2, row c, as outlined above. If discrepancies exist between the other person's ratings and your self-image, how can you reconcile them? It may be that the other person does not know you very well and has not sampled enough of your behavior to make an accurate judgment; on the other hand, you may not have communicated what you are really like to him, or your self-image may be in error.

4. *Awareness of Self-Presentation (Forms 1 and 3).* Are you aware that another person may see you differently from the way you feel about yourself? Compare your self-ratings on Form 1 with your predictions of O's ratings on Form 3. Explain any differences. If there are differences, have you intentionally presented yourself differently than you are? One student who recently completed this exercise explained, "I know I am able to present myself in a better light than I actually am."

5. Mutuality (Forms 2 and 5). This comparison will let you know the extent to which you and your partner have mutual feelings about each other. If your partner is a good friend, it is likely that there will be close correspondence between your ratings of each other (and also between your self-ratings Forms 1 and 4). Byrne (1961), Newcomb (1961), and others have demonstrated that friends are often more similar in their opinions and traits than people who do not befriend each other.

Optional Further Discussion

Relationships Among Relationships. A glance at Fig. 1 reveals some interesting properties of the relationships discussed above. If any two sides of a triangle in the diagram are known, then the third side is determined. For example, if you are perfectly accurate (a, Forms 3 and 5) and perfectly congruent (b, Forms 2 and 3), you and your partner's ratings will be perfectly mutual. To prove this to yourself, try it with the other possible triangle of relationships that can be drawn with the data from Data Sheet 2 (self-presentation, awareness of self-presentation, and accuracy). For example, if you are perfectly accurate (a), but are aware that your self-presentation is discrepant (d), then actual self-presentation (c) will be discrepant by exactly the same amount as (d). When neither of the two known sides of a person's perception triangle represent perfect agreement (that is, they both reveal some discrepancy), it is necessary to know the direction of the two discrepancies in order to deduce the third relationship. For example, if the accuracy difference (a) is 1 and the congruency difference (b) is also 1, mutuality can either be perfect or the discrepancy will be 2, depending on whether accuracy and congruency are discrepant in the same direction or whether they cancel each other in determining mutuality. Think about explanations for the various relationships. For instance, you may be aware that you are capable of hiding your faults or good qualities by presenting yourself inaccurately.

Components of Rating Variance. In person perception research we are usually more interested in how the qualities of the person being rated (O) are perceived than we are in P's rating habits and biases. A general problem, then, is to separate out that portion of the ratings that reflects only the rater's style. This is not a simple task (the problems of identifying and calculating the components of rating variance are discussed in detail by Gage and Cronbach 1955), and it requires that we set up a situation where (1) many raters (Ps) rate the same person (O), and (2) each rater (P) rates many similar people (Os). Then "within" comparisons made by the same rater can be contrasted with "between" ratings made by different raters (this procedure is described by Dornbusch et al, 1965).

From this kind of analysis it is possible to make corrections for style biases. For example, a constant factor could be added or subtracted to a person's ratings if he was generally more positive or critical than other raters; especially if the level (high or low) of his ratings was not related to his ability to discriminate among O's.

Self-Presentation. As you may have already found out, raters (including yourself) may find it difficult to make frank appraisals in a nonanonymous situation such as exists in this exercise. In a society that emphasizes tact and

good manners on one hand and modesty and constraint on the other, both negative and positive evaluations may be suppressed, that is, raters may want to protect O's feelings by tempering negative evaluations, but they also may be reluctant to express positive ratings when these ratings will be seen by O. It may be possible to encourage more direct, honest ratings by using an anonymous procedure. Can you think of other means of overcoming this problem?

References

Byrne, D. Interpersonal attraction as a function of affiliation need and attitude similarity. *Journal of Abnormal and Social Psychology* 1961, **62**,713–715.

Subjects were asked to express their feelings about "strangers" (actually the subjects received faked attitude scales filled out by the experimenter) whose views on a number of topics were either similar or dissimilar to their own attitudes. Subjects rated strangers whose attitudes were similar as more likeable and more desirable as a work partner. In addition, similar strangers were estimated to be more intelligent, better adjusted, and better informed.

Dornbusch, S. M., Hastorf, A. H., Richardson, S. A., Muzzy, R. E., and Vreeland, R. S. The perceiver and the perceived: Their relative influence on the categories of interpersonal perception. *Journal of Personality and Social Psychology,* 1965, **1**, 434–440.

Nine- to eleven-year-old camp children were asked to describe each other; then their free descriptions were coded according to the categories they used. There was more category overlap when one child described two others than when two children described the same other child. The results are interpreted as emphasizing the importance of the describer's cognitive structure.

Gage, N. L., and Cronbach, L. J. Conceptual and methodological problems in interpersonal perception. *Psychological Review,* 1955, **62**, 411–422.

This somewhat technical article breaks down simple person perception ratings into several useful components. For example, the component "elevation" simply tells us which part of the rating scale is used, and is not relevant to the discriminations a rater makes between the people he rates.

Heider, F. *The psychology of interpersonal relations.* New York: Wiley, 1958.

In this book, Heider develops and analyzes the common sense or naive psychology of perception.

Newcomb, T. M. *The acquaintance process.* New York: Holt, Rinehart and Winston, 1961.

In this classic field study, 17 male students who had never previously met were invited to live together for one year in a University of Michigan cooperative house. Newcomb's book described the changes that took place during the year as the men became acquaint-

ed. In general, those men who shared common values and beliefs came to prefer each other as friends.

Tagiuri, R. Person perception. In G. Lindzey and E. Aronson (Eds.), *The handbook of social psychology,* rev. ed., Vol. 3, Reading, Mass.: Addison-Wesley, 1969 395–449.

The most recent and comprehensive survey of the person perception literature.

Tagiuri, R., Bruner, J. S., and Blake, R. R. On the relation between feelings and the perception of feelings among members of small groups. In Maccoby, Newcomb, and Hartley (Eds.), *Readings in social psychology*, 1958, pp. 110–116.

Sociometric data were gathered from a wide variety of groups including sailors, campers, and seminar members. Accuracy, mutuality, and congruency (see Fig. 1, p. 358) relationships computed from these data are compared and discussed.

Data Sheet 1
General Positive-Negative Rating Scores

(a) P's Ratings

Scale item	Form 1 P's self rating	Form 3 P's estimate of O's rating of P	Form 5 O's rating of P
A			
B			
C			
D			
TOTAL			
Mean = $\frac{\text{Total}}{4}$			

(b) O's Ratings

Scale item	Form 4 O's self rating	Form 6 O's estimate of P's rating of O	Form 2 P's rating of O
A			
B			
C			
D			
TOTAL			
Mean = $\frac{\text{Total}}{4}$			

Data Sheet 2
Comparisons between Forms for P

(a) RAW SCORES

FORM NUMBER	SCALE ITEM			
	E	F	G	H
Form 1 (P → P)				
Form 2 (P → O)				
Form 3 P(O → P)				
Form 5 (O → P)				

(b) DIFFERENCES
(Use positive numbers only)

COMPARISONS	SCALE ITEM				TOTAL DISCREPANCY (Add across items)
	E	F	G	H	
a.* 3 & 5 (Accuracy, sensitivity)					
b. 2 & 3 (Congruency)					
c. 1 & 5 (Self-presentation)					
d. 1 & 3 (Awareness of self-presentation)					
e. 2 & 5 (Mutuality)					

*Letters a, b, c, d, and e refer to Figure 1, p. 358.

Person Rating Form____Name_______________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong __X__ _____ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ _____ __X__ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong _____ __X__ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ __X__ _____ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong _____ _____ __X__ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ __X__ _____ _____ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong _____ _____ _____ __X__ _____ _____ _____ Weak

Use the scales below to describe ______________________________. Please be as accurate as possible.

A FAIR _____ _____ _____ _____ _____ _____ _____ UNFAIR

B DISHONEST _____ _____ _____ _____ _____ _____ _____ HONEST

C KIND _____ _____ _____ _____ _____ _____ _____ CRUEL

D UNPLEASANT _____ _____ _____ _____ _____ _____ _____ PLEASANT

E TALKATIVE _____ _____ _____ _____ _____ _____ _____ QUIET

F UNUSUAL _____ _____ _____ _____ _____ _____ _____ AVERAGE

G SERIOUS _____ _____ _____ _____ _____ _____ _____ HAPPY-GO-LUCKY

H INDEPENDENT _____ _____ _____ _____ _____ _____ _____ SOCIABLE

Person Rating Form___Name____________________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong _X_ ___ ___ ___ ___ ___ ___ Weak
Strong ___ ___ ___ ___ ___ ___ _X_ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong ___ _X_ ___ ___ ___ ___ ___ Weak
Strong ___ ___ ___ ___ ___ _X_ ___ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong ___ ___ _X_ ___ ___ ___ ___ Weak
Strong ___ ___ ___ ___ _X_ ___ ___ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong ___ ___ ___ _X_ ___ ___ ___ Weak

Use the scales below to describe ____________________. Please be as accurate as possible.

A FAIR ___ ___ ___ ___ ___ ___ ___ UNFAIR

B DISHONEST ___ ___ ___ ___ ___ ___ ___ HONEST

C KIND ___ ___ ___ ___ ___ ___ ___ CRUEL

D UNPLEASANT ___ ___ ___ ___ ___ ___ ___ PLEASANT

E TALKATIVE ___ ___ ___ ___ ___ ___ ___ QUIET

F UNUSUAL ___ ___ ___ ___ ___ ___ ___ AVERAGE

G SERIOUS ___ ___ ___ ___ ___ ___ ___ HAPPY-GO-LUCKY

H INDEPENDENT ___ ___ ___ ___ ___ ___ ___ SOCIABLE

Person Rating Form ____ Name ____________________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong __X__ _____ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ _____ __X__ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong _____ __X__ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ __X__ _____ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong _____ _____ __X__ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ __X__ _____ _____ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong _____ _____ _____ __X__ _____ _____ _____ Weak

Use the scales below to describe ______________________________. Please be as accurate as possible.

A FAIR _____ _____ _____ _____ _____ _____ _____ UNFAIR

B DISHONEST _____ _____ _____ _____ _____ _____ _____ HONEST

C KIND _____ _____ _____ _____ _____ _____ _____ CRUEL

D UNPLEASANT _____ _____ _____ _____ _____ _____ _____ PLEASANT

E TALKATIVE _____ _____ _____ _____ _____ _____ _____ QUIET

F UNUSUAL _____ _____ _____ _____ _____ _____ _____ AVERAGE

G SERIOUS _____ _____ _____ _____ _____ _____ _____ HAPPY-GO-LUCKY

H INDEPENDENT _____ _____ _____ _____ _____ _____ _____ SOCIABLE

Person Rating Form____Name________________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong __X__ _____ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ _____ __X__ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong _____ __X__ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ __X__ _____ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong _____ _____ __X__ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ __X__ _____ _____ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong _____ _____ _____ __X__ _____ _____ _____ Weak

Use the scales below to describe ____________________________. Please be as accurate as possible.

A FAIR _____ _____ _____ _____ _____ _____ _____ UNFAIR

B DISHONEST _____ _____ _____ _____ _____ _____ _____ HONEST

C KIND _____ _____ _____ _____ _____ _____ _____ CRUEL

D UNPLEASANT _____ _____ _____ _____ _____ _____ _____ PLEASANT

E TALKATIVE _____ _____ _____ _____ _____ _____ _____ QUIET

F UNUSUAL _____ _____ _____ _____ _____ _____ _____ AVERAGE

G SERIOUS _____ _____ _____ _____ _____ _____ _____ HAPPY-GO-LUCKY

H INDEPENDENT _____ _____ _____ _____ _____ _____ _____ SOCIABLE

Person Rating Form____Name________________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong __X__ _____ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ _____ __X__ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong _____ __X__ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ __X__ _____ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong _____ _____ __X__ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ __X__ _____ _____ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong _____ _____ _____ __X__ _____ _____ _____ Weak

Use the scales below to describe ______________________________. Please be as accurate as possible.

A FAIR _____ _____ _____ _____ _____ _____ _____ UNFAIR

B DISHONEST _____ _____ _____ _____ _____ _____ _____ HONEST

C KIND _____ _____ _____ _____ _____ _____ _____ CRUEL

D UNPLEASANT _____ _____ _____ _____ _____ _____ _____ PLEASANT

E TALKATIVE _____ _____ _____ _____ _____ _____ _____ QUIET

F UNUSUAL _____ _____ _____ _____ _____ _____ _____ AVERAGE

G SERIOUS _____ _____ _____ _____ _____ _____ _____ HAPPY-GO-LUCKY

H INDEPENDENT _____ _____ _____ _____ _____ _____ _____ SOCIABLE

Person Rating Form___Name____________________

Instructions:

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) extremely well, place your "X" as follows:

Strong __X__ _____ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ _____ __X__ Weak

If you feel that the adjective at one end of the scale describes (the person you are rating) (you) quite well (but not extremely well), you should place your "X" as follows:

Strong _____ __X__ _____ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ _____ __X__ _____ Weak

If you feel that the adjective at one end of the scale only slightly describes (the person you are rating) (you), you should place your "X" as follows:

Strong _____ _____ __X__ _____ _____ _____ _____ Weak
Strong _____ _____ _____ _____ __X__ _____ _____ Weak

If you consider (the person you are rating) (you) to be neutral on the scale, or if the scale is completely irrelevant, then you should place your "X" in the middle space.

Strong _____ _____ _____ __X__ _____ _____ _____ Weak

Use the scales below to describe ______________________________. Please be as accurate as possible.

A FAIR _____ _____ _____ _____ _____ _____ _____ UNFAIR

B DISHONEST _____ _____ _____ _____ _____ _____ _____ HONEST

C KIND _____ _____ _____ _____ _____ _____ _____ CRUEL

D UNPLEASANT _____ _____ _____ _____ _____ _____ _____ PLEASANT

E TALKATIVE _____ _____ _____ _____ _____ _____ _____ QUIET

F UNUSUAL _____ _____ _____ _____ _____ _____ _____ AVERAGE

G SERIOUS _____ _____ _____ _____ _____ _____ _____ HAPPY-GO-LUCKY

H INDEPENDENT _____ _____ _____ _____ _____ _____ _____ SOCIABLE

Exercise 4

Generalized Attitude Change

Unlike previous exercises, the present one does not outline a specific experimental procedure. Instead, the following material is intended to provide a broad context within which the student can design his own experiment in the area of attitude change. Perhaps more so than any other area in social psychology, experiments on attitude change have relied on the "pretest-posttest" and "posttest-only" design. In the pretest-posttest design, subjects are first given a pretest on a particular attitude issue, then they are exposed to some form of persuasive communication, and finally they are given a posttest. Control subjects are merely given the pretest and posttest without a persuasive communication. Any difference between experimental and comparison groups is attributed to the persuasive communication.

In the posttest-only design, none of the subjects are given a pretest. Half of the subjects are randomly assigned to receive the persuasive communication and then the posttest; the other half of the subjects receive the posttest only. Any difference between the group that received only the posttest and the group that received both posttest and persuasive communication is attributed to the persuasive communication. Exercise 6 on communicator attractiveness-similarity and Exercise 5 on role playing are both examples of posttest-only designs. Furthermore, both of those experiments contain at least the rudiments of a persuasive communication. Both of those exercises also contain posttest material; and Exercise 2 on attitude measurement describes how you can develop pre- and posttesting materials on their own.

With these materials, you are in a position to replicate almost all of the studies on attitude change in the literature. We shall not provide any specific experimental materials in this exercise; instead, we intend this exercise to be

the most unstructured in the book. It is up to you to devise an experiment on attitude change.

Experimental manipulations in the area of attitude change have traditionally been divided into three areas: (1) manipulations relevant to the source of the communication; for example, communicator attractiveness-similarity, (2) manipulations relevant to the content of the message itself; for example, order of arguments, emotional versus rational appeals, and (3) variables relevant to the audience in the communication; for example, educated versus uneducated, high versus low self-esteem. Excellent reviews of this literature are available elsewhere (Insko, 1967; McGuire, 1968; Kiesler, Collins, and Miller, 1969; see also the latest *Annual Review of Psychology*). A quick reading of any of these sources should suggest several experiments.

Communicator Characteristics. The remainder of this introduction does not attempt to reproduce the literature reviews of the attitude change literature cited above. Instead, we shall present a brief overview of some of the theoretical and empirical controversies in the area of communicator characteristics. A somewhat more detailed presentation using the same outline is presented in Collins (1970). Communicator characteristic experiments are usually relatively simple—the subjects are led to think one thing about the speaker in one condition and another in the other condition. Thus this area should be a good one in which to design your first experiment.

Most studies examining communicator credibility have used gross, unanalytic experimental manipulations. Half of the subjects would be told that the communication was from the Archangel Gabriel and the other half told that Mephistopheles, or some other personification of evil itself, was the source of the communication. All subjects read or listen to exactly the same persuasive communication. The typical finding is that there is more attitude change when the communication is attributed to a positive source than when it is attributed to a negative source—even though both groups of subjects listen to or read exactly the same persuasive communication. However, relatively few investigators have gotten past their excitement in demonstrating this basic fact and have moved on to the question: "Why should some kinds of communicators produce more attitude change than others?"

In their important book, *Communication and Persuasion*, Hovland, Janis, and Kelley (1953) suggest two important components of communicator effectiveness: *trustworthiness* and *expertness*. In a typical study, a communication on atomic submarines was attributed either to J. Robert Oppenheimer or to *Pravda*. Thus they used the typical communicator effectiveness paradigm in which a communication is attributed to some famous person who is presumed to be trustworthy and expert or to another famous communicator, who is presumed to be untrustworthy and unexpert. Thus trustworthiness and expertness are confounded (confounding is discussed in the Introduc-

tion, pages 17-22). Hovland, Janis, and Kelley's general conclusion was that expertness was an important variable; but they were able to find little evidence in support of the trustworthiness hypothesis.

Another variable that may affect perceptions of the communicator is the role of forewarning of persuasive intent. Presumably, a communicator who announces that he intends to persuade his audience should be seen as less trustworthy than a communicator who has—or at least has not announced—an explicit personal interest in manipulating his audience. A typical experimental procedure was first reported by Allyn and Festinger (1961). Some subjects were told that the experimenter would ask their opinions at the end of a persuasive talk (opinion orientation-forewarned condition). Other subjects were told that they would be asked to rate the speaker's personality (personality orientation-unforewarned condition). In another experiment, subjects were warned by a footnote which stated that the article was intended to change attitudes (Kiesler and Kiesler, 1964). Although the results are not overwhelming, there is some evidence that forewarning inhibits attitude change. (See also Freedman and Sears, 1965; Apsler and Sears, 1968.)

Other investigators have reported that overheard persuasive communications are sometimes more effective than persuasive communications when the speaker knows he has an audience. If this is so, it would support the trustworthiness hypothesis. Presumably, a speaker cannot deliberately attempt to manipulate an audience if he does not know he has an audience. Thus a speaker might be seen as more trustworthy if the audience thinks they are eavesdropping on his comment.

However, several experiments report that overheard communications are more effective only if the topic is involving and socially desirable for the audience (Walster and Festinger, 1962; Brock and Becker, 1965). The area of overheard persuasive communications is a fascinating one to research—both because there is some controversy as to the effectiveness of overheard communications and because, even if true, we yet do not understand why overheard communication should be effective only for topics which are involving to the audience.

In Exercise 6, we have already reviewed some experiments demonstrating that an audience is more likely to be persuaded by a communicator that it sees as similar to itself than a communicator that it sees as dissimilar. In one fascinating study (Berscheid, 1966), however, subjects showed more attitude change only when the dimensions of similarity were relevant to the persuasive communication. Thus this area also requires further research. Is it true that only relevant similarities produce attitude change? Why should an audience be more persuaded by a similar communicator than a nonsimilar communicator?

Another interesting area much in need of research is the legitimacy of the

communicator. Are there certain kinds of people (parents, teachers, policeman) who have the right to tell us what to think? In other words, are there certain people who have the authority of legitimacy to demand attitude change? Although there are a number of theories of social influence (for example, French and Raven, 1959) that stress legitimacy or authority as a source of power, there are relatively few experiments that directly study this phenomenon (see Raven and French, 1958a, 1958b, for two interesting experiments on this problem). The area of legitimate social influence may be an interesting one in which to study the relationship between attitude and behavior. Do many audiences feel that legitimate sources or authoritative sources have the right to prescribe certain behaviors but do not have the right to prescribe certain attitudes?

Speaker-Audience Discrepancy. There is at least one other controversy in the area of source characteristics and attitude change that is worth mentioning. It is generally accepted that there is a curvilinear relationship between attitude change and the discrepancy between the audience's position and the speaker's position. As the speaker and audience begin to diverge in their points of view, the audience's attitude change first increases, and then begins to decrease. In other words, it is possible for a speaker to advocate so discrepant a position that the audience shows no attitude change at all. Some investigators have reported that communicators with high prestige or credibility can successfully advocate more discrepant messages than sources with low prestige or credibility. Just why this should be the case is an interesting theoretical question. As of this printing, Aronson, Turner, and Carlsmith (1963), Bergin (1962), Bochner and Insko (1966), Johnson (1966), and Koslin, Stoops, and Loh (1967) were among the more important journal articles on this complex controversy. The controversy is also reviewed in both McGuire (1968) and Kiesler, Collins, and Miller (1969). Since it is a current and controversial topic, it is also probably discussed in the most recent chapter on attitude change in the *Annual Review of Psychology*.

The above issues by no means exhaust the possible topics for study in the areas of attitude change. Perhaps because the general structure of the pretest-posttest and posttest-only design make it so easy to design a new experiment in attitude change, that area may contain more published articles than any other in the last ten years of social psychology. Whether the student attempts to follow up on one of the areas briefly reviewed above, elaborate on an experiment discussed in some social psychology textbook, or strikes out with a hypothesis of his own, it should be easy for him to design and execute an original study on attitude change.

References

Allyn, J., and Festinger, L. The effectiveness of unanticipated persuasive communications. *Journal of Abnormal and Social Psychology,* 1961, **62**, 35–40.

Apsler, R., and Sears, D. O. Warning, personal involvement, and attitude change. *Journal of Personality and Social Psychology,* 1968, **9**, 162–166.

Aronson, E., Turner, J. A., and Carlsmith, J. M. Communicator credibility and communication discrepancy as determinants of opinion change. *Journal of Abnormal and Social Psychology,* 1963, **67**, 31–36.

Bergin, A. R. The effect of dissonant persuasive communications upon changes in a self-referring attitude. *Journal of Personality,* 1962, **30**, 423–438.

Berscheid, E. Opinion change and communicator-communicatee similarity and dissimilarity. *Journal of Personality and Social Psychology,* 1966, **4**, 670–680.

Bochner, S., and Insko, C. A. Communicator discrepancy, source credibility, and opinion change. *Journal of Personality and Social Psychology,* 1966, **4**, 614–621.

Brock, T. C., and Becker, L. A. Ineffectiveness of "overheard" counterpropaganda, *Journal of Personality and Social Psychology,* 1965, **2**, 654–660.

Collins, B. E. *Social psychology: Social influence, attitude change, small groups, and prejudice.* Reading, Mass.: Addison-Wesley, 1970.

Freedman, J. L., and Sears, D. O. Warning, distraction, and resistance to influence. *Journal of Personality and Social Psychology,* 1965, **1**, 262–266.

French, J. R. P., Ross, I. C., Kirby, S., Nelson, J. R., and Smyth, P. Employee participation in a program of industrial change. *Personnel*, 1958, 19–29.

Hovland, C. I., Janis, I. L., and Kelley, H. H. *Communication and persuasion*. New Haven: Yale University Press, 1953.

Insko, C. A. *Theories of attitude change.* New York: Appleton-Century-Crofts, 1967.

Johnson, H. H. Some effects of discrepancy level on responses to negative information about one's self. *Sociometry*, 1966, **29**, 52–65.

Kiesler, C. A., Collins, B. E., and Miller, N. *Attitude change: A critical analysis of theoretical approaches.* New York: John Wiley, 1969.

Kiesler, C. A., and Kiesler, S. B. Role of forewarning in persuasive communications. *Journal of Abnormal Social Psychology,* 1964, **68**, 547–549.

Koslin, B. L., Stoops, J. W., and Loh, W. D. Source characteristics and communication discrepancy as determinants of attitude change and conformity. *Journal of Experimental Social Psychology,* 1967, **3**, 230–242.

McGuire, W. J. The nature of attitudes and attitude change. In G. Lindzey and E. Aronson (Eds.), *Handbook of social psychology,* second edition, Vol. 3, Reading, Mass: Addison-Wesley, 1968. Pages 136–314.

Raven, B. H., and French, J. R. P., Jr. Group support, legitimate power, and social influence. *Journal of Personality,* 1958a, **26**, 400–409.

Raven, B. H., and French, J. R. P., Jr. Legitimate power, coercive power, and observability in social influence. *Sociometry,* 1958b, **21**, 83–97.

Walster, E., and Festinger, L. The effectiveness of "overheard" persuasive communications. *Journal of Abnormal and Social Psychology,* 1962, **65**, 395–402.

Calculation of t-Test for Difference Between Means of Two Independent Random Samples Measured on the Same Variable

Step 1. Enter the data (1 number per subject) for one group or condition in Column I.

Step 2. Enter the data (1 number per subject) for the other group or condition in Column III.

Step 3. Square each number in Column I and enter the result in Column II.

Step 4. Square each number in Column III and enter the result in Column IV.

	Column I	Column II	Column III	Column IV
Subject Number	Scores on Group 1	(Column I)^2	Scores on Group 2	(Column II)^2
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
TOTALS	to Box C	to Box D	to Box E	to Box F

Step 5. Add all the numbers in each column and enter the totals in Boxes C, D, E & F.

Step 6. Perform the arithmetic computations indicated to the right of each of Boxes A to W.

Step 7. The far right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

	Arithmetic Operations	Statistical Explanation
A	Count the number of subjects in Column I.	n_1
B	Count the number of subjects in Column III.	n_2
C	Enter the total of Column I.	
D	Enter the total of Column II.	
E	Enter the total of Column III.	
F	Enter the total of Column IV.	
G	Add Boxes A + B.	Total $N = (n_1 + n_2)$
H	Add Boxes C + E.	Total sum = ΣX
I	Square number in Box H.	$(\Sigma X)^2$
J	Add Boxes D + F.	Total sum of squares = ΣX^2
K	Multiply Box G by Box J.	$(N) \times (\Sigma X^2)$ or $N\Sigma X^2$
L	Subtract 1 from Box G.	$N - 1$
M	Multiply Box G by Box L.	$(N) \times (N - 1)$ or $N(N - 1)$
N	Box K minus Box I.	$N\Sigma X^2 - (\Sigma X)^2$
O	Divide Box N by Box M.	Estimate of $S^2 = \dfrac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$
P	Divide Box C by Box A.	Mean of Group 1 $(\overline{X}_1)$
Q	Divide Box E by Box B.	Mean of Group 2 $(\overline{X}_2)$
R	Subtract Box Q from Box P (Ignore + or − sign; enter as positive number).	Mean Difference $\lvert \overline{X}_1 - \overline{X}_2 \rvert$
S	Multiply Box A by Box B.	$(n_1) \times (n_2)$ or $n_1 n_2$
T	Divide Box G by Box S.	$\dfrac{n_1 + n_2}{n_1 n_2}$
U	Multiply Box T by Box O.	$S^2 \dfrac{n_1 + n_2}{n_1 n_2}$

V ☐ Take the square root of Box U (from the square root table, p. 409). $\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}$

W ☐ Divide Box R by Box V. $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$

Step 8. Use the table below to evaluate the t-value in Box W. The term *"df"* (degrees of freedom) equals the total number of subjects (Box G) minus 2. If the t statistic you computed in Box W is larger than the value in the table below, these results (the difference between Groups 1 and 2) would have occurred by chance less than 5 times in 100.

Distribution of t for $p = .05$[1]

df	*t(.05)*
2	4.30
4	2.78
6	2.46
8	2.31
10	2.23
15	2.13
20	2.09
25	2.06
30	2.04
60	2.00
120	1.98
∞	1.96

[1] Abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* published by Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

The values of p are given for a two-tailed test.

Calculation of t-Test for Difference Between Means of Two Independent Random Samples Measured on the Same Variable

Step 1. Enter the data (1 number per subject) for one group or condition in Column I.
Step 2. Enter the data (1 number per subject) for the other group or condition in Column III.
Step 3. Square each number in Column I and enter the result in Column II.
Step 4. Square each number in Column III and enter the result in Column IV.

	Column I	Column II	Column III	Column IV
Subject Number	Scores on Group 1	$(\text{Column I})^2$	Scores on Group 2	$(\text{Column II})^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
TOTALS	to Box C	to Box D	to Box E	to Box F

Step 5. Add all the numbers in each column and enter the totals in Boxes C, D, E & F.

Step 6. Perform the arithmetic computations indicated to the right of each of Boxes A to W.

Step 7. The far right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

Box	Arithmetic Operations	Statistical Explanation
A	Count the number of subjects in Column I.	n_1
B	Count the number of subjects in Column III.	n_2
C	Enter the total of Column I.	
D	Enter the total of Column II.	
E	Enter the total of Column III.	
F	Enter the total of Column IV.	
G	Add Boxes A + B.	Total $N = (n_1 + n_2)$
H	Add Boxes C + E.	Total sum = ΣX
I	Square number in Box H.	$(\Sigma X)^2$
J	Add Boxes D + F.	Total sum of squares = ΣX^2
K	Multiply Box G by Box J.	$(N) \times (\Sigma X^2)$ or $N\Sigma X^2$
L	Subtract 1 from Box G.	$N - 1$
M	Multiply Box G by Box L.	$(N) \times (N - 1)$ or $N(N - 1)$
N	Box K minus Box I.	$N\Sigma X^2 - (\Sigma X)^2$
O	Divide Box N by Box M.	Estimate of $S^2 = \dfrac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$
P	Divide Box C by Box A.	Mean of Group 1 $(\overline{X}_1)$
Q	Divide Box E by Box B.	Mean of Group 2 $(\overline{X}_2)$
R	Subtract Box Q from Box P (Ignore + or − sign; enter as positive number).	Mean Difference $\lvert \overline{X}_1 - \overline{X}_2 \rvert$
S	Multiply Box A by Box B.	$(n_1) \times (n_2)$ or $n_1 n_2$
T	Divide Box G by Box S.	$\dfrac{n_1 + n_2}{n_1 n_2}$
U	Multiply Box T by Box O.	$S^2 \dfrac{n_1 + n_2}{n_1 n_2}$

V [ ] Take the square root of Box U (from the square root table, p. 409). $\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}$

W [ ] Divide Box R by Box V. $t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$

Step 8. Use the table below to evaluate the t-value in Box W. The term *"df"* (degrees of freedom) equals the total number of subjects (Box G) minus 2. If the t statistic you computed in Box W is larger than the value in the table below, these results (the difference between Groups 1 and 2) would have occurred by chance less than 5 times in 100.

Distribution of t for $p = .05$[1]

df	*t(.05)*
2	4.30
4	2.78
6	2.46
8	2.31
10	2.23
15	2.13
20	2.09
25	2.06
30	2.04
60	2.00
120	1.98
∞	1.96

[1] Abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* published by Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

The values of p are given for a two-tailed test.

Calculation of t-Test for Difference Between Means of Two Independent Random Samples Measured on the Same Variable

Step 1. Enter the data (1 number per subject) for one group or condition in Column I.
Step 2. Enter the data (1 number per subject) for the other group or condition in Column III.
Step 3. Square each number in Column I and enter the result in Column II.
Step 4. Square each number in Column III and enter the result in Column IV.

	Column I	Column II	Column III	Column IV
Subject Number	Scores on Group 1	(Column I)^2	Scores on Group 2	(Column II)^2
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
TOTALS	to Box C	to Box D	to Box E	to Box F

Step 5. Add all the numbers in each column and enter the totals in Boxes C, D, E & F.

Step 6. Perform the arithmetic computations indicated to the right of each of Boxes A to W.

Step 7. The far right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

	Arithmetic Operations	Statistical Explanation
A ☐	Count the number of subjects in Column I.	n_1
B ☐	Count the number of subjects in Column III.	n_2
C ☐	Enter the total of Column I.	
D ☐	Enter the total of Column II.	
E ☐	Enter the total of Column III.	
F ☐	Enter the total of Column IV.	
G ☐	Add Boxes A + B.	Total $N = (n_1 + n_2)$
H ☐	Add Boxes C + E.	Total sum = ΣX
I ☐	Square number in Box H.	$(\Sigma X)^2$
J ☐	Add Boxes D + F.	Total sum of squares = ΣX^2
K ☐	Multiply Box G by Box J.	$(N) \times (\Sigma X^2)$ or $N\Sigma X^2$
L ☐	Subtract 1 from Box G.	$N - 1$
M ☐	Multiply Box G by Box L.	$(N) \times (N - 1)$ or $N(N - 1)$
N ☐	Box K minus Box I.	$N\Sigma X^2 - (\Sigma X)^2$
O ☐	Divide Box N by Box M.	Estimate of $S^2 = \dfrac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$
P ☐	Divide Box C by Box A.	Mean of Group 1 $(\overline{X}_1)$
Q ☐	Divide Box E by Box B.	Mean of Group 2 $(\overline{X}_2)$
R ☐	Subtract Box Q from Box P (Ignore + or − sign; enter as positive number).	Mean Difference $\lvert\overline{X}_1 - \overline{X}_2\rvert$
S ☐	Multiply Box A by Box B.	$(n_1) \times (n_2)$ or $n_1 n_2$
T ☐	Divide Box G by Box S.	$\dfrac{n_1 + n_2}{n_1 n_2}$
U ☐	Multiply Box T by Box O.	$S^2 \dfrac{n_1 + n_2}{n_1 n_2}$

V [ ] Take the square root of Box U (from the square root table, p. 409). $\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}$

W [ ] Divide Box R by Box V. $t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$

Step 8. Use the table below to evaluate the t-value in Box W. The term *"df"* (degrees of freedom) equals the total number of subjects (Box G) minus 2. If the t statistic you computed in Box W is larger than the value in the table below, these results (the difference between Groups 1 and 2) would have occurred by chance less than 5 times in 100.

Distribution of t for $p = .05$[1]

df	*t(.05)*
2	4.30
4	2.78
6	2.46
8	2.31
10	2.23
15	2.13
20	2.09
25	2.06
30	2.04
60	2.00
120	1.98
∞	1.96

[1] Abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* published by Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

The values of p are given for a two-tailed test.

Calculation of t-Test for Difference Between Means of Two Independent Random Samples Measured on the Same Variable

Step 1. Enter the data (1 number per subject) for one group or condition in Column I.

Step 2. Enter the data (1 number per subject) for the other group or condition in Column III.

Step 3. Square each number in Column I and enter the result in Column II.

Step 4. Square each number in Column III and enter the result in Column IV.

	Column I	Column II	Column III	Column IV
Subject Number	Scores on Group 1	$(\text{Column I})^2$	Scores on Group 2	$(\text{Column II})^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
TOTALS	to Box C	to Box D	to Box E	to Box F

Step 5. Add all the numbers in each column and enter the totals in Boxes C, D, E & F.

Step 6. Perform the arithmetic computations indicated to the right of each of Boxes A to W.

Step 7. The far right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

Box	**Arithmetic Operations**	**Statistical Explanation**
A	Count the number of subjects in Column I.	n_1
B	Count the number of subjects in Column III.	n_2
C	Enter the total of Column I.	
D	Enter the total of Column II.	
E	Enter the total of Column III.	
F	Enter the total of Column IV.	
G	Add Boxes A + B.	Total $N = (n_1 + n_2)$
H	Add Boxes C + E.	Total sum = ΣX
I	Square number in Box H.	$(\Sigma X)^2$
J	Add Boxes D + F.	Total sum of squares = ΣX^2
K	Multiply Box G by Box J.	$(N) \times (\Sigma X^2)$ or $N\Sigma X^2$
L	Subtract 1 from Box G.	$N - 1$
M	Multiply Box G by Box L.	$(N) \times (N - 1)$ or $N(N - 1)$
N	Box K minus Box I.	$N\Sigma X^2 - (\Sigma X)^2$
O	Divide Box N by Box M.	Estimate of $S^2 = \dfrac{N\Sigma X^2 - (\Sigma X)^2}{N(N-1)}$
P	Divide Box C by Box A.	Mean of Group 1 $(\overline{X}_1)$
Q	Divide Box E by Box B.	Mean of Group 2 $(\overline{X}_2)$
R	Subtract Box Q from Box P (Ignore + or – sign; enter as positive number).	Mean Difference $\lvert\overline{X}_1 - \overline{X}_2\rvert$
S	Multiply Box A by Box B.	$(n_1) \times (n_2)$ or $n_1 n_2$
T	Divide Box G by Box S.	$\dfrac{n_1 + n_2}{n_1 n_2}$
U	Multiply Box T by Box O.	$S^2 \dfrac{n_1 + n_2}{n_1 n_2}$

V ☐ Take the square root of Box U (from the square root table, p. 409). $\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}$

W ☐ Divide Box R by Box V. $t = \dfrac{\overline{X_1} - \overline{X_2}}{\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$

Step 8. Use the table below to evaluate the t-value in Box W. The term *"df"* (degrees of freedom) equals the total number of subjects (Box G) minus 2. If the t statistic you computed in Box W is larger than the value in the table below, these results (the difference between Groups 1 and 2) would have occurred by chance less than 5 times in 100.

Distribution of t for $p = .05$[1]

df	*t(.05)*
2	4.30
4	2.78
6	2.46
8	2.31
10	2.23
15	2.13
20	2.09
25	2.06
30	2.04
60	2.00
120	1.98
∞	1.96

[1] Abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* published by Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

The values of p are given for a two-tailed test.

Calculation of t-Test for Difference Between Means of Two Independent Random Samples Measured on the Same Variable

Step 1. Enter the data (1 number per subject) for one group or condition in Column I.
Step 2. Enter the data (1 number per subject) for the other group or condition in Column III.
Step 3. Square each number in Column I and enter the result in Column II.
Step 4. Square each number in Column III and enter the result in Column IV.

	Column I	Column II	Column III	Column IV
Subject Number	Scores on Group 1	$(\text{Column I})^2$	Scores on Group 2	$(\text{Column II})^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
TOTALS	to Box C	to Box D	to Box E	to Box F

Step 5. Add all the numbers in each column and enter the totals in Boxes C, D, E & F.

Step 6. Perform the arithmetic computations indicated to the right of each of Boxes A to W.

Step 7. The far right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

	Arithmetic Operations	Statistical Explanation
A ☐	Count the number of subjects in Column I.	n_1
B ☐	Count the number of subjects in Column III.	n_2
C ☐	Enter the total of Column I.	
D ☐	Enter the total of Column II.	
E ☐	Enter the total of Column III.	
F ☐	Enter the total of Column IV.	
G ☐	Add Boxes A + B.	Total $N = (n_1 + n_2)$
H ☐	Add Boxes C + E.	Total sum = ΣX
I ☐	Square number in Box H.	$(\Sigma X)^2$
J ☐	Add Boxes D + F.	Total sum of squares = ΣX^2
K ☐	Multiply Box G by Box J.	$(N) \times (\Sigma X^2)$ or $N\Sigma X^2$
L ☐	Subtract 1 from Box G.	$N - 1$
M ☐	Multiply Box G by Box L.	$(N) \times (N - 1)$ or $N(N - 1)$
N ☐	Box K minus Box I.	$N\Sigma X^2 - (\Sigma X)^2$
O ☐	Divide Box N by Box M.	Estimate of $S^2 = \dfrac{N\Sigma X^2 - (\Sigma X)^2}{N(N - 1)}$
P ☐	Divide Box C by Box A.	Mean of Group 1 $(\overline{X}_1)$
Q ☐	Divide Box E by Box B.	Mean of Group 2 $(\overline{X}_2)$
R ☐	Subtract Box Q from Box P (Ignore + or − sign; enter as positive number).	Mean Difference $\lvert \overline{X}_1 - \overline{X}_2 \rvert$
S ☐	Multiply Box A by Box B.	$(n_1) \times (n_2)$ or $n_1 n_2$
T ☐	Divide Box G by Box S.	$\dfrac{n_1 + n_2}{n_1 n_2}$
U ☐	Multiply Box T by Box O.	$S^2 \dfrac{n_1 + n_2}{n_1 n_2}$

V ☐ Take the square root of Box U (from the square root table, p. 409). $\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}$

W ☐ Divide Box R by Box V. $t = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{S^2\left(\frac{n_1 + n_2}{n_1 n_2}\right)}}$

Step 8. Use the table below to evaluate the t-value in Box W. The term *"df"* (degrees of freedom) equals the total number of subjects (Box G) minus 2. If the t statistic you computed in Box W is larger than the value in the table below, these results (the difference between Groups 1 and 2) would have occurred by chance less than 5 times in 100.

Distribution of t for $p = .05$[1]

df	*t(.05)*
2	4.30
4	2.78
6	2.46
8	2.31
10	2.23
15	2.13
20	2.09
25	2.06
30	2.04
60	2.00
120	1.98
∞	1.96

[1] Abridged from Table III of Fisher & Yates: *Statistical Tables for Biological, Agricultural and Medical Research,* published by Oliver & Boyd Ltd., Edinburgh, by permission of the author and publishers.

The values of p are given for a two-tailed test.

Square Root Table

Box U	$\sqrt{\ }$	Box U	$\sqrt{\ }$	Box U	$\sqrt{\ }$	Box U	$\sqrt{\ }$	Box U	$\sqrt{\ }$	Box U	$\sqrt{\ }$
0.1	0.32	3.2	1.79	6.4	2.53	9.5	3.08	36.0	6.00	68.0	8.25
0.2	0.45	3.3	1.82	6.5	2.55	9.6	3.10	37.0	6.08	69.0	8.31
0.3	0.55	3.4	1.84	6.6	2.57	9.7	3.11	38.0	6.16	70.0	8.37
0.4	0.63	3.5	1.87	6.7	2.59	9.8	3.13	39.0	6.24	71.0	8.43
0.5	0.71	3.6	1.90	6.8	2.61	9.9	3.15	40.0	6.32	72.0	8.48
0.6	0.77	3.7	1.92	6.9	2.62	10.0	3.16	41.0	6.40	73.0	8.54
0.7	0.84	3.8	1.95	7.0	2.64	11.0	3.32	42.0	6.48	74.0	8.60
0.8	0.89	3.9	1.97	7.1	2.66	12.0	3.46	43.0	6.56	75.0	8.66
0.9	0.94	4.0	2.00	7.2	2.68	13.0	3.60	44.0	6.63	76.0	8.72
1.0	1.00	4.1	2.02	7.3	2.70	14.0	3.74	45.0	6.71	77.0	8.77
1.1	1.05	4.2	2.05	7.4	2.72	15.0	3.87	46.0	6.78	78.0	8.83
1.2	1.10	4.3	2.07	7.5	2.74	16.0	4.00	47.0	6.85	79.0	8.89
1.3	1.14	4.4	2.10	7.6	2.76	17.0	4.12	48.0	6.93	80.0	8.94
1.4	1.18	4.5	2.12	7.7	2.77	18.0	4.24	49.0	7.00	81.0	9.00
1.5	1.22	4.6	2.14	7.8	2.79	19.0	4.36	50.0	7.07	82.0	9.05
1.6	1.26	4.7	2.17	7.9	2.81	20.0	4.47	51.0	7.14	83.0	9.11
1.7	1.30	4.8	2.19	8.0	2.83	21.0	4.58	52.0	7.21	84.0	9.16
1.8	1.34	4.9	2.21	8.1	2.85	22.0	4.69	53.0	7.28	85.0	9.22
1.9	1.38	5.0	2.23	8.2	2.86	23.0	4.79	54.0	7.35	86.0	9.27
2.0	1.41	5.1	2.26	8.3	2.88	24.0	4.90	55.0	7.42	87.0	9.33
2.1	1.45	5.2	2.28	8.4	2.90	25.0	5.00	56.0	7.48	88.0	9.38
2.2	1.48	5.3	2.30	8.5	2.91	26.0	5.10	57.0	7.55	89.0	9.43
2.3	1.52	5.4	2.32	8.6	2.93	27.0	5.20	58.0	7.61	90.0	9.49
2.4	1.55	5.5	2.34	8.7	2.95	28.0	5.29	59.0	7.68	91.0	9.54
2.5	1.58	5.6	2.37	8.8	2.97	29.0	5.38	60.0	7.74	92.0	9.59
2.6	1.61	5.7	2.39	8.9	2.98	30.0	5.48	61.0	7.81	93.0	9.64
2.7	1.64	5.8	2.41	9.0	3.00	31.0	5.57	62.0	7.87	94.0	9.69
2.8	1.67	5.9	2.43	9.1	3.02	32.0	5.66	63.0	7.94	95.0	9.75
2.9	1.70	6.0	2.45	9.2	3.03	33.0	5.74	64.0	8.00	96.0	9.80
3.0	1.73	6.1	2.47	9.3	3.05	34.0	5.83	65.0	8.06	97.0	9.85
3.1	1.76	6.2	2.49	9.4	3.06	35.0	5.92	66.0	8.12	98.0	9.90
		6.3	2.51					67.0	8.18	99.0	9.95

Calculation of Chi Square (χ^2)

The starting point for a 2 × 2 chi square (the only kind discussed here) is a 2 × 2 table shown below which divides two groups of subjects on some variable *X*.

	These people are in Group I (e.g., experimental, males, witnessed model, etc.)	These people are in Group II (e.g., control, females did not witness model, etc.)
Number of people who have high value on variable *X* (e.g., conformed to model, characteristic present, etc.)	Box A	Box B
Number of people who have low value on variable *X* (e.g., did not conform, characteristic not present, etc.)	Box C	Box D

When an Exercise calls for a χ^2 computation, the appropriate 2 × 2 table is described in the Exercise.

Step 1. Check to see that the sum of Box A + B + C + D = the total number of subjects (*N*) that participated in the study.

Step 2. Perform the arithmetic computations indicated to the right of Boxes E through O on the reverse side.

Box Letter		Arithmetic Operation
E		Add Box A + Box B
F		Add Box C + Box D
G		Add Box A + Box C
H		Add Box B + Box D
I		Multiply Box A by Box D
J		Multiply Box B by Box C
K		Subtract Box J from Box I
L		Square Box K
M		Multiply Box K by Box N (the total number of Ss)
N		Multiply Box E by Box F
O		Multiply the number in Box N above by Box G
P		Multiply Box O by Box H
Q		Divide Box M by Box P

Step 3. If the χ^2 value in Box Q is larger than 3.84, these results would have occurred by chance less than 6 times in 100 ($p < .05$).

Computational formula for $2 \times 2\ \chi^2$:

$$\chi^{2\,1} = \frac{N(AD - BC)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

[1] This simple computational formula is uncorrected for continuity.

Appendix B, #2

Calculation of Chi Square (χ^2)

The starting point for a 2 $\times$ 2 chi square (the only kind discussed here) is a 2 $\times$ 2 table shown below which divides two groups of subjects on some variable *X*.

	These people are in Group I (e.g., experimental, males, witnessed model, etc.)	These people are in Group II (e.g., control, females did not witness model; etc.)
Number of people who have high value on variable *X* (e.g., conformed to model, characteristic present, etc.)	Box A	Box B
Number of people who have low value on variable *X* (e.g., did not conform, characteristic not present, etc.)	Box C	Box D

When an Exercise calls for a χ^2 computation, the appropriate 2 $\times$ 2 table is described in the Exercise.

Step 1. Check to see that the sum of Box A + B + C + D = the total number of subjects (*N*) that participated in the study.

Step 2. Perform the arithmetic computations indicated to the right of Boxes E through O on the reverse side.

Box Letter		Arithmetic Operation
E		Add Box A + Box B
F		Add Box C + Box D
G		Add Box A + Box C
H		Add Box B + Box D
I		Multiply Box A by Box D
J		Multiply Box B by Box C
K		Subtract Box J from Box I
L		Square Box K
M		Multiply Box K by Box N (the total number of Ss)
N		Multiply Box E by Box F
O		Multiply the number in Box N above by Box G
P		Multiply Box O by Box H
Q		Divide Box M by Box P

Step 3. If the χ^2 value in Box Q is larger than 3.84, these results would have occurred by chance less than 6 times in 100 ($p < .05$).

Computational formula for 2×2 χ^2:

χ^2[1] $$= \frac{N(AD - BC)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

[1] This simple computational formula is uncorrected for continuity.

Calculation of Chi Square (χ^2)

The starting point for a 2 × 2 chi square (the only kind discussed here) is a 2 × 2 table shown below which divides two groups of subjects on some variable *X*.

	These people are in Group I (e.g., experimental, males, witnessed model, etc.)	These people are in Group II (e.g., control, females did not witness model, etc.)
Number of people who have high value on variable *X* (e.g., conformed to model, characteristic present, etc.)	Box A	Box B
Number of people who have low value on variable *X* (e.g., did not conform, characteristic not present, etc.)	Box C	Box D

When an Exercise calls for a χ^2 computation, the appropriate 2 × 2 table is described in the Exercise.

Step 1. Check to see that the sum of Box A + B + C + D = the total number of subjects (*N*) that participated in the study.

Step 2. Perform the arithmetic computations indicated to the right of Boxes E through O on the reverse side.

Box Letter		Arithmetic Operation
E		Add Box A + Box B
F		Add Box C + Box D
G		Add Box A + Box C
H		Add Box B + Box D
I		Multiply Box A by Box D
J		Multiply Box B by Box C
K		Subtract Box J from Box I
L		Square Box K
M		Multiply Box K by Box N (the total number of Ss)
N		Multiply Box E by Box F
O		Multiply the number in Box N above by Box G
P		Multiply Box O by Box H
Q		Divide Box M by Box P

Step 3. If the χ^2 value in Box Q is larger than 3.84, these results would have occurred by chance less than 6 times in 100 ($p < .05$).

Computational formula for 2 X 2 χ^2:

$$\chi^{2\,1} = \frac{N(AD - BC)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

[1] This simple computational formula is uncorrected for continuity.

Calculation of Chi Square (χ^2)

The starting point for a 2 × 2 chi square (the only kind discussed here) is a 2 × 2 table shown below which divides two groups of subjects on some variable X.

	These people are in Group I (e.g., experimental, males, witnessed model, etc.)	These people are in Group II (e.g., control, females did not witness model, etc.)
Number of people who have high value on variable X (e.g., conformed to model, characteristic present, etc.)	Box A	Box B
Number of people who have low value on variable X (e.g., did not conform, characteristic not present, etc.)	Box C	Box D

When an Exercise calls for a χ^2 computation, the appropriate 2 × 2 table is described in the Exercise.

Step 1. Check to see that the sum of Box A + B + C + D = the total number of subjects (N) that participated in the study.

Step 2. Perform the arithmetic computations indicated to the right of Boxes E through O on the reverse side.

Box Letter		Arithmetic Operation
E		Add Box A + Box B
F		Add Box C + Box D
G		Add Box A + Box C
H		Add Box B + Box D
I		Multiply Box A by Box D
J		Multiply Box B by Box C
K		Subtract Box J from Box I
L		Square Box K
M		Multiply Box K by Box N (the total number of Ss)
N		Multiply Box E by Box F
O		Multiply the number in Box N above by Box G
P		Multiply Box O by Box H
Q		Divide Box M by Box P

Step 3. If the χ^2 value in Box Q is larger than 3.84, these results would ha
occurred by chance less than 6 times in 100 ($p < .05$).

Computational formula for 2 X 2 χ^2:

$$\chi^{2}\,[1] = \frac{N(AD - BC)^2}{(A+B)\ (C+D)\ (A+C)\ (B+D)}$$

[1] This simple computational formula is uncorrected for continuity.

Calculation of Rank-Order Correlation Coefficient (Rho)

(For some exercises, data sheets for steps 1 to 4 are provided with the exercise. If so, use the exercise data sheet and start with step 5.)

Step 1. Fill in rank order for both variable 1 and variable 2. Do not enter subject score. The subject who has the highest score on variable 1 gets a "1"; the second highest gets a "2." The subject with the lowest score gets the lowest rank (15 for 15 subjects, 28 for 28 subjects, etc.). Subjects in a tie are assigned the average rank for which they are tied. For example, subjects tied for 5 and 6 both get a rank of 5.5 (5 + 6 = 11; 11 ÷ 2 = 5.5). Subjects tied for 5, 6, and 7 all get a rank of 6 (5 + 6 + 7 = 18; 18 ÷ 3 = 6). When ranks are tied, the calculations described here for rho are not exact, but they are a fair approximation.

Step 2. Subtract the numbers in Column II from those in Column I. Enter the "difference" in Column III. As a check on your work, the sum of Column III should be 0.

Step 3. Square each difference score and enter this "(difference score)^2" in Column IV.

Step 4. Add all of the values of the (difference score)^2 in Column IV and enter this sum in "Box A" at the bottom of Column IV.

Step 5. Perform the arithmetic computations indicated to the right of Boxes A to G.

Step 6. The right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

Step 7. Use the rho table to evaluate the rho value in Box G. If the rho statistic you computed in Box G is larger than the value in the table, your results would have occurred by chance less than 5 times in 100.

	Column I	Column II	Column III	Column IV
Subject Number	Rank on Variable I	Rank on Variable 2	Difference	$(\text{Difference})^2$
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				

Check	
$\boxed{}$	$\boxed{}$
Sum of Column III should = 0	Box A Sum of $(\text{differences})^2$ Σd^2

	Arithmetic Operations	Statistical Explanation
A	Copy number from Box A at bottom of Column IV.	Σd^2
B	Multiply Box A by 6.	$(6)(\Sigma d^2)$
C	Count the number of subjects and enter the number in Box C (Box C = number of entries in columns I).	N
D	Cube the number in Box C [Box C X Box C X Box C].	N^3
E	Subtract Box C from Box D.	$N^3 - N$
F	Divide Box B by Box E.	$\dfrac{(6)(\Sigma d^2)}{N^3 - N}$
G	Subtract Box F from 1.	Rho $(\rho) = 1 - \left[\dfrac{(6)\ (\Sigma d^2)}{N^3 - N}\right]$

Table of Critical Values of Rho,
the Spearman Rank Correlation Coefficient for ρ = .05[1]

N Box C	.05 Significance level (one-tailed test)
4	1.000
6	.829
8	.643
10	.564
12	.506
14	.456
16	.425
18	.399
20	.377
22	.359
24	.343
26	.329
28	.317
30	.306

[1] Adapted and abridged from Olds, E. G. 1938. Distributions of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Statist.,* **9**, 133-148, and from Olds, E. G. 1949. The 5% significance levels for sums of squares of rank differences and a correction. *Ann. Math. Statist.,* **20**, 117-118.

Calculation of Rank-Order Correlation Coefficient (Rho)

(For some exercises, data sheets for steps 1 to 4 are provided with the exercise. If so, use the exercise data sheet and start with step 5.)

Step 1. Fill in rank order for both variable 1 and variable 2. Do not enter subject score. The subject who has the highest score on variable 1 gets a "1"; the second highest gets a "2." The subject with the lowest score gets the lowest rank (15 for 15 subjects, 28 for 28 subjects, etc.). Subjects in a tie are assigned the average rank for which they are tied. For example, subjects tied for 5 and 6 both get a rank of 5.5 ($5 + 6 = 11$; $11 \div 2 = 5.5$). Subjects tied for 5, 6, and 7 all get a rank of 6 ($5 + 6 + 7 = 18$; $18 \div 3 = 6$). When ranks are tied, the calculations described here for rho are not exact, but they are a fair approximation.

Step 2. Subtract the numbers in Column II from those in Column I. Enter the "difference" in Column III. As a check on your work, the sum of Column III should be 0.

Step 3. Square each difference score and enter this "(difference score)^2" in Column IV.

Step 4. Add all of the values of the (difference score)^2 in Column IV and enter this sum in "Box A" at the bottom of Column IV.

Step 5. Perform the arithmetic computations indicated to the right of Boxes A to G.

Step 6. The right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

Step 7. Use the rho table to evaluate the rho value in Box G. If the rho statistic you computed in Box G is larger than the value in the table, your results would have occurred by chance less than 5 times in 100.

	Column I	Column II	Column III	Column IV
Subject Number	Rank on Variable I	Rank on Variable 2	Difference	(Difference)^2
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				

Check

Sum of Column III should = 0	Box A Sum of (differences)^2 Σd^2

	Arithmetic Operations	Statistical Explanation
A [ ]	Copy number from Box A at bottom of Column IV.	Σd^2
B [ ]	Multiply Box A by 6.	$(6)(\Sigma d^2)$
C [ ]	Count the number of subjects and enter the number in Box C (Box C = number of entries in columns I).	N
D [ ]	Cube the number in Box C [Box C X Box C X Box C].	N^3
E [ ]	Subtract Box C from Box D.	$N^3 - N$
F [ ]	Divide Box B by Box E.	$\frac{(6)(\Sigma d^2)}{N^3 - N}$
G [ ]	Subtract Box F from 1.	Rho $(\rho) = 1 - \left[\frac{(6)(\Sigma d^2)}{N^3 - N}\right]$

Table of Critical Values of Rho,
the Spearman Rank Correlation Coefficient for $\rho = .05$[1]

N Box C	.05 Significance level (one-tailed test)
4	1.000
6	.829
8	.643
10	.564
12	.506
14	.456
16	.425
18	.399
20	.377
22	.359
24	.343
26	.329
28	.317
30	.306

[1] Adapted and abridged from Olds, E. G. 1938. Distributions of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Statist.*, **9**, 133-148, and from Olds, E. G. 1949. The 5% significance levels for sums of squares of rank differences and a correction. *Ann. Math. Statist.*, **20**, 117-118.

Calculation of Rank-Order Correlation Coefficient (Rho)

(For some exercises, data sheets for steps 1 to 4 are provided with the exercise. If so, use the exercise data sheet and start with step 5.)

Step 1. Fill in rank order for both variable 1 and variable 2. Do not enter subject score. The subject who has the highest score on variable 1 gets a "1"; the second highest gets a "2." The subject with the lowest score gets the lowest rank (15 for 15 subjects, 28 for 28 subjects, etc.). Subjects in a tie are assigned the average rank for which they are tied. For example, subjects tied for 5 and 6 both get a rank of 5.5 ($5 + 6 = 11$; $11 \div 2 = 5.5$). Subjects tied for 5, 6, and 7 all get a rank of 6 ($5 + 6 + 7 = 18$; $18 \div 3 = 6$). When ranks are tied, the calculations described here for rho are not exact, but they are a fair approximation.

Step 2. Subtract the numbers in Column II from those in Column I. Enter the "difference" in Column III. As a check on your work, the sum of Column III should be 0.

Step 3. Square each difference score and enter this "(difference score)^2" in Column IV.

Step 4. Add all of the values of the (difference score)^2 in Column IV and enter this sum in "Box A" at the bottom of Column IV.

Step 5. Perform the arithmetic computations indicated to the right of Boxes A to G.

Step 6. The right-hand column labeled "Statistical Explanation" is included for the benefit of more advanced students. It explains the operations performed in the left-hand column using statistical notation.

Step 7. Use the rho table to evaluate the rho value in Box G. If the rho statistic you computed in Box G is larger than the value in the table, your results would have occurred by chance less than 5 times in 100.

	Column I	Column II	Column III	Column IV
Subject Number	Rank on Variable I	Rank on Variable 2	Difference	(Difference)^2
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				

Check	
Sum of Column III should = 0	Box A Sum of (differences)^2 Σd^2

	Arithmetic Operations	Statistical Explanation
A ☐	Copy number from Box A at bottom of Column IV.	Σd^2
B ☐	Multiply Box A by 6.	$(6)(\Sigma d^2)$
C ☐	Count the number of subjects and enter the number in Box C (Box C = number of entries in columns I).	N
D ☐	Cube the number in Box C [Box C × Box C × Box C].	N^3
E ☐	Subtract Box C from Box D.	$N^3 - N$
F ☐	Divide Box B by Box E.	$\frac{(6)(\Sigma d^2)}{N^3 - N}$
G ☐	Subtract Box F from 1.	Rho $(\rho) = 1 - \left[\frac{(6)(\Sigma d^2)}{N^3 - N}\right]$

Table of Critical Values of Rho,
the Spearman Rank Correlation Coefficient for $\rho = .05$[1]

N Box C	.05 Significance level (one-tailed test)
4	1.000
6	.829
8	.643
10	.564
12	.506
14	.456
16	.425
18	.399
20	.377
22	.359
24	.343
26	.329
28	.317
30	.306

[1] Adapted and abridged from Olds, E. G. 1938. Distributions of sums of squares of rank differences for small numbers of individuals. *Ann. Math. Statist.,* **9**, 133-148, and from Olds, E. G. 1949. The 5% significance levels for sums of squares of rank differences and a correction. *Ann. Math. Statist.,* **20**, 117-118.

Appendix D

Ethical Standards of Psychologists

The psychologist believes in the dignity and worth of the individual human being. He is committed to increasing man's understanding of himself and others. While pursuing this endeavor, he protects the welfare of any person who may seek his service or of any subject, human or animal, that may be the object of his study. He does not use his professional position or relationships, nor does he knowingly permit his own services to be used by others, for purposes inconsistent with these values. While demanding for himself freedom of inquiry and communication, he accepts the responsibility this freedom confers: for competence where he claims it, for objectivity in the report of his findings, and for consideration of the best interests of his colleagues and of society.

SPECIFIC PRINCIPLES

PRINCIPLE 1. RESPONSIBILITY

The psychologist,[1] committed to increasing man's understanding of man, places high value on objectivity and integrity, and maintains the highest standards in the services he offers.

a. As a scientist, the psychologist believes that society will be best served when he investigates where his judgment indicates investigation is needed; he plans his research in such a way as to minimize the possibility that his findings will be misleading; and he publishes full reports of his work, never discarding without explanation data which may modify the interpretation of results.

b. As a teacher, the psychologist recognizes his primary obligation to help others acquire knowledge and skill, and to maintain high standards of scholarship.

c. As a practitioner, the psychologist knows that he bears a heavy social responsibility because his work may touch intimately the lives of others.

SOURCE: *American Psychologist*, 1963, XVIII (1), 56-60.

[1] A student of psychology who assumes the role of psychologist shall be considered a psychologist for the purpose of this code of ethics. Principles 1a, 6, and 16 are most relevant for experimentation.

PRINCIPLE 2. COMPETENCE

The maintenance of high standards of professional competence is a responsibility shared by all psychologists, in the interest of the public and of the profession as a whole.

a. Psychologists discourage the practice of psychology by unqualified persons and assist the public in identifying psychologists competent to give dependable professional service. When a psychologist or a person identifying himself as a psychologist violates ethical standards, psychologists who know firsthand of such activities attempt to rectify the situation. When such a situation cannot be dealt with informally, it is called to the attention of the appropriate local, state, or national committee on professional ethics, standards and practices.

b. The psychologist recognizes the boundaries of his competence and the limitations of his techniques and does not offer services or use techniques that fail to meet professional standards established in particular fields. The psychologist who engages in practice assists his client in obtaining professional help for all important aspects of his problem that fall outside the boundaries of his own competence. This principle requires, for example, that provision be made for the diagnosis and treatment of relevant medical problems and for referral to or consultation with other specialists.

c. The psychologist in clinical work recognizes that his effectiveness depends in good part upon his ability to maintain sound interpersonal relations, that temporary or more enduring aberrations in his own personality may interfere with this ability or distort his appraisals of others. There he refrains from undertaking any activity in which his personal problems are likely to result in inferior professional services or harm to a client; or, if he is already engaged in such an activity when he becomes aware of his personal problems, he seeks competent professional assistance to determine whether he should continue or terminate his services to his client.

PRINCIPLE 3. MORAL AND LEGAL STANDARDS

The psychologist in the practice of his profession shows sensible regard for the social codes and moral expectations of the community in which he works, recognizing that violations of accepted moral and legal standards on his part may involve his clients, students, or colleagues in damaging personal conflicts, and impugn his own name and the reputation of his profession.

PRINCIPLE 4. MISREPRESENTATION

The psychologist avoids misrepresentation of his own professional qualifications, affiliations, and purposes, and those of the institutions and organizations with which he is associated.

a. A psychologist does not claim either directly or by implication professional qualifications that differ from his actual qualifications, nor does he misrepresent his affiliation with any institution, organization, or individual, nor lead others to assume he has affiliations that he does not have. The psychologist is responsible for correcting others who misrepresent his professional qualifications or affiliations.

b. The psychologist does not misrepresent an institution or organization with which he is affiliated by ascribing to it characteristics that it does not have.

c. A psychologist does not use his affiliation with the American Psychological Association or its Divisions for purposes that are not consonant with the stated purposes of the Association.

d. A psychologist does not associate himself with or permit his name to be used in connection with any services or products in such a way as to misrepresent them, the degree of his responsibility for them, or the nature of his affiliation.

PRINCIPLE 5. PUBLIC STATEMENTS

Modesty, scientific caution, and due regard for the limits of present knowledge characterize all statements of psychologists who supply information to the public, either directly or indirectly.

a. Psychologists who interpret the science of psychology or the services of psychologists to clients or to the general public have an obligation to report fairly and accurately. Exaggeration, sensationalism, superficiality, and other kinds of misrepresentation are avoided.

b. When information about psychological procedures and techniques is given, care is taken to indicate that they should be used only by persons adequately trained in their use.

c. A psychologist who engages in radio or television activities does not participate in commercial announcements recommending purchase or use of a product.

PRINCIPLE 6. CONFIDENTIALITY

Safeguarding information about an individual that has been obtained by the psychologist in the course of his teaching, practice, or investigation is a primary obligation of the psychologist. Such information is not communicated to others unless certain important conditions are met.

a. Information received in confidence is revealed only after most careful deliberation and when there is clear and imminent danger to an individual or to society, and then only to appropriate professional workers or public authorities.

b. Information obtained in clinical or consulting relationships, or evaluative data concerning children, students, employees, and others are discussed only for professional purposes and only with persons clearly concerned with the case. Written and oral reports should present only data germane to the purposes of the evaluation; every effort should be made to avoid undue invasion of privacy.

c. Clinical and other case materials are used in classroom teaching and writing only when the identity of the persons involved is adequately disguised.

d. The confidentiality of professional communications about individuals is maintained. Only when the originator and other persons involved give their express permission is a confidential professional communication shown to the individual concerned. The psychologist is responsible for informing the client of the limits of the confidentiality.

e. Only after explicit permission has been granted is the identity of research subjects published. When data have been published without permission for identification, the psychologist assumes responsibility for adequately disguising their sources.

f. The psychologist makes provision for the maintenance of confidentiality in the preservation and ultimate disposition of confidential records.

PRINCIPLE 7. CLIENT WELFARE

The psychologist respects the integrity and protects the welfare of the person or group with whom he is working.

a. The psychologist in industry, education, and other situations in which conflicts of interest may arise among various parties, as between management and labor, or between the client and employer of the psychologist, defines for himself the nature and direction of his loyalties and responsibilities and keeps all parties concerned informed of these commitments.

b. When there is a conflict among professional workers, the psychologist is concerned primarily with the welfare of any client involved and only secondarily with the interest of his own professional group.

c. The psychologist attempts to terminate a clinical or consulting relationship when it is reasonably clear to the psychologist that the client is not benefiting from it.

d. The psychologist who asks that an individual reveal personal information in the course of interviewing, testing, or evaluation, or who allows such information to be divulged to him, does so only after making certain that the responsible person is fully aware of the purposes of the interview, testing, or evaluation and of the ways in which the information may be used.

e. In cases involving referral, the responsibility of the psychologist for the

welfare of the client continues until this responsibility is assumed by the professional person to whom the client is referred or until the relationship with the psychologist making the referral has been terminated by mutual agreement. In situations where referral, consultation, or other changes in the conditions of the treatment are indicated and the client refuses referral, the psychologist carefully weighs the possible harm to the client, to himself, and to his profession that might ensue from continuing the relationship.

f. The psychologist who requires the taking of psychological tests for didactic, classification, or research purposes protects the examinees by insuring that the tests and test results are used in a professional manner.

g. When potentially disturbing subject matter is presented to students, it is discussed objectively, and efforts are made to handle constructively any difficulties that arise.

h. Care must be taken to insure an appropriate setting for clinical work to protect both client and psychologist from actual or imputed harm and the profession from censure.

PRINCIPLE 8. CLIENT RELATIONSHIP

The psychologist informs his prospective client of the important aspects of the potential relationship that might affect the client's decision to enter the relationship.

a. Aspects of the relationship likely to affect the client's decision include the recording of an interview, the use of interview material for training purposes, and observation of an interview by other persons.

b. When the client is not competent to evaluate the situation (as in the case of a child), the person responsible for the client is informed of the circumstances which may influence the relationship.

c. The psychologist does not normally enter into a professional relationship with members of his own family, intimate friends, close associates, or others whose welfare might be jeopardized by such a dual relationship.

PRINCIPLE 9. IMPERSONAL SERVICES

Psychological services for the purpose of diagnosis, treatment, or personalized advice are provided only in the context of a professional relationship, and are not given by means of public lectures or demonstrations, newspaper or magazine articles, radio or television programs, mail, or similar media.

a. The preparation of personnel reports and recommendations based on test data secured solely by mail is unethical unless such appraisals are an integral part of a continuing client relationship with a company, as a result of which the consulting psychologist has intimate knowledge of the client's

personnel situation and can be assured thereby that his written appraisals will be adequate to the purpose and will be properly interpreted by the client. These reports must not be embellished with such detailed analyses of the subject's personality traits as would be appropriate only after intensive interviews with the subject. The reports must not make specific recommendations as to employment or placement of the subject which go beyond the psychologist's knowledge of the job requirements of the company. The reports must not purport to eliminate the company's need to carry on such other regular employment or personnel practices as appraisal of the work history, checking of references, past performance in the company.

PRINCIPLE 10. ANNOUNCEMENT OF SERVICES

A psychologist adheres to professional rather than commercial standards in making known his availability for professional services.

a. A psychologist does not directly solicit clients for individual diagnosis or therapy.

b. Individual listings in telephone directories are limited to name, highest relevant degree, certification status, address, and telephone number. They may also include identification in a few words of the psychologist's major areas of practice; for example, child therapy, personnel selection, industrial psychology. Agency listings are equally modest.

c. Announcements of individual private practice are limited to a simple statement of the name, highest relevant degree, certification or diplomate status, address, telephone number, office hours, and a brief explanation of the types of services rendered. Announcements of agencies may list names of staff members with their qualifications. They conform in other particulars with the same standards as individual announcements, making certain that the true nature of the organization is apparent.

d. A psychologist or agency announcing nonclinical professional services may use brochures that are descriptive of services rendered but not evaluative. They may be sent to professional persons, schools, business firms, government agencies, and other similar organizations.

e. The use in a brochure of "testimonials from satisfied users" is unacceptable. The offer of a free trial of services is unacceptable if it operates to misrepresent in any way the nature or the efficacy of the services rendered by the psychologist. Claims that a psychologist has unique skills or unique devices not available to others in the profession are made only if the special efficacy of these unique skills or devices has been demonstrated by scientifically acceptable evidence.

f. The psychologist must not encourage (nor, within his power, even

allow) a client to have exaggerated ideas as to the efficacy of services rendered. Claims made to clients about the efficacy of his services must not go beyond those which the psychologist would be willing to subject to professional scrutiny through publishing his results and his claims in a professional journal.

PRINCIPLE 11. INTERPROFESSIONAL RELATIONS

A psychologist acts with integrity in regard to colleagues in psychology and in other professions.

a. A psychologist does not normally offer professional services to a person receiving psychological assistance from another professional worker except by agreement with the other worker or after the termination of the client's relationship with the other professional worker.

b. The welfare of clients and colleagues requires that psychologists in joint practice or corporate activities make an orderly and explicit arrangement regarding the conditions of their association and its possible termination. Psychologists who serve as employers of other psychologists have an obligation to make similar appropriate arrangements.

PRINCIPLE 12. REMUNERATION

Financial arrangements in professional practice are in accord with professional standards that safeguard the best interest of the client and the profession.

a. In establishing rates for professional services, the psychologist considers carefully both the ability of the client to meet the financial burden and the charges made by other professional persons engaged in comparable work. He is willing to contribute a portion of his services to work for which he receives little or no financial return.

b. No commission or rebate or any other form of remuneration is given or received for referral of clients for professional services.

c. The psychologist in clinical or counseling practice does not use his relationship with clients to promote, for personal gain or the profit of an agency, commercial enterprises of any kind.

d. A psychologist does not accept a private fee or any other form of remuneration for professional work with a person who is entitled to his services through an institution or agency. The policies of a particular agency may make explicit provision for private work with its clients by members of its staff, and in such instances the client must be fully apprised of all policies affecting him.

PRINCIPLE 13. TEST SECURITY

Psychological tests and other assessment devices, the value of which depends in part on the naivete of the subject, are not reproduced or described in popular publications in ways that might invalidate the techniques. Access to such devices is limited to persons with professional interests who will safeguard their use.

a. Sample items made up to resemble those of tests being discussed may be reproduced in popular articles and elsewhere, but scorable tests and actual test items are not reproduced except in professional publications.

b. The psychologist is responsible for the control of psychological tests and other devices and procedures used for instruction when their value might be damaged by revealing to the general public their specific contents or underlying principles.

PRINCIPLE 14. TEST INTERPRETATION

Test scores, like test materials, are released only to persons who are qualified to interpret and use them properly.

a. Materials for reporting test scores to parents, or which are designed for self-appraisal purposes in schools, social agencies, or industry are closely supervised by qualified psychologists or counselors with provisions for referring and counseling individuals when needed.

b. Test results or other assessment data used for evaluation or classification are communicated to employers, relatives, or other appropriate persons in such a manner as to guard against misinterpretation or misuse. In the usual case, an interpretation of the test result rather than the score is communicated.

c. When test results are communicated directly to parents and students, they are accompanied by adequate interpretive aids or advice.

PRINCIPLE 15. TEST PUBLICATION

Psychological tests are offered for commercial publication only to publishers who present their tests in a professional way and distribute them only to qualified users.

a. A test manual, technical handbook, or other suitable report on the test is provided which describes the method of constructing and standardizing the test, and summarizes the validation research.

b. The populations for which the test has been developed and the purposes for which it is recommended are stated in the manual. Limitations upon the test's dependability, and aspects of its validity on which research is lacking or incomplete, are clearly stated. In particular, the manual contains a

warning regarding interpretations likely to be made which have not yet been substantiated by research.

c. The catalog and manual indicate the training or professional qualifications required for sound interpretation of the test.

d. The test manual and supporting documents take into account the principles enunciated in the *Technical Recommendations for Psychological Tests and Diagnostic Techniques.*

e. Test advertisements are factual and descriptive rather than emotional and persuasive.

PRINCIPLE 16. RESEARCH PRECAUTIONS

The psychologist assumes obligations for the welfare of his research subjects, both animal and human.

a. Only when a problem is of scientific significance and it is not practicable to investigate it in any other way is the psychologist justified in exposing research subjects, whether children or adults, to physical or emotional stress as part of an investigation.

b. When a reasonable possibility of injurious aftereffects exists, research is conducted only when the subjects or their responsible agents are fully informed of this possibility and agree to participate nevertheless.

c. The psychologist seriously considers the possibility of harmful aftereffects and avoids them, or removes them as soon as permitted by the design of the experiment.

d. A psychologist using animals in research adheres to the provisions of the Rules Regarding Animals, drawn up by the Committee on Precautions and Standards in Animal Experimentation and adopted by the American Psychological Association.

PRINCIPLE 17. PUBLICATION CREDIT

Credit is assigned to those who have contributed to a publication, in proportion to their contribution, and only to these.

a. Major contributions of a professional character, made by several persons to a common project, are recognized by joint authorship. The experimenter or author who has made the principal contribution to a publication is identified as the first listed.

b. Minor contributions of a professional character, extensive clerical or similar nonprofessional assistance, and other minor contributions are acknowledged in footnotes or in an introductory statement.

c. Acknowledgment through specific citations is made for unpublished as well as published material that has directly influenced the research or writing.

d. A psychologist who compiles and edits for publication the contributions of others publishes the symposium or report under the title of the committee or symposium, with his own name appearing as chairman or editor among those of the other contributors or committee members.

PRINCIPLE 18. RESPONSIBILITY TOWARD ORGANIZATION

A psychologist respects the rights and reputation of the institute or organization with which he is associated.

a. Materials prepared by a psychologist as a part of his regular work under specific direction of his organization are the property of that organization. Such materials are released for use or publication by a psychologist in accordance with policies of authorization, assignment of credit, and related matters which have been established by his organization.

b. Other material resulting incidentally from activity supported by any agency, and for which the psychologist rightly assumes individual responsibility, is published with disclaimer for any responsibility on the part of the supporting agency.

PRINCIPLE 19. PROMOTIONAL ACTIVITIES

The psychologist associated with the development or promotion of psychological devices, books, or other products offered for commercial sale is responsible for ensuring that such devices, books, or products are presented in a professional and factual way.

a. Claims regarding performance, benefits, or results are supported by scientifically acceptable evidence.

b. The psychologist does not use professional journals for the commercial exploitation of psychological products, and the psychologist-editor guards against such misuse.

c. The psychologist with a financial interest in the sale or use of a psychological product is sensitive to possible conflict of interest in his promotion of such products and avoids compromise of his professional responsibilities and objectives.

Example and Brief Discussion of Written Research Reports

JAMES P. PSYCHOLOGIST
Northcentral State University

This paper briefly describes and illustrates a format for writing up experimental reports (American Psychological Association, 1969). Each report begins with a brief abstract of 100–120 words, which should include statements of the (1) general problem, (2) method, (3) important results, and (4) main conclusion(s). It is desirable for the abstract to include number and type of subjects, a summary of the research design, and the significance level of results. The introduction or problem section briefly describes the theoretical background of the problem and *states* the hypotheses. The next section describes the method and procedures used to create independent variables and measure dependent variables. Then the results are presented. The final section, the discussion, covers the methodological and theoretical implications of the study.

INTRODUCTION OR PROBLEM (This heading is not necessary)

Each of the exercises in this book contains a brief introduction which sets the exercise in a methodological and theoretical context. Using this material and any additional reading, the author describes the theoretical and methodological problems that are relevant in a particular exercise. Why should this particular problem be studied? What are the hypotheses? How were they derived? What is the theoretical or empirical basis for expecting that the hypotheses will work out? Why is this particular procedure a good way to study this problem?

METHOD

Overview of Design

In this section you will want to briefly describe each of the experimental and comparison or control groups that you used in the procedure. In one or two paragraphs you should be able to give the reader a quick overview of the manipulation of the independent variable and the measurement of the dependent variable (see Introduction page 14).

Subjects

In the method section, the author should offer a precise description of the subjects used in that exercise. What was their approximate age? Were they

females or males? Where did you recruit them? What were the criteria for including and excluding subjects? Were there any other distinguishing characteristics of the subjects used that the reader should know about?

Procedure

In this section you will provide a detailed, chronological account of exactly how the experiment was conducted, that is, a step-by-step account of what the experimenter did in order to gather the data. It will include verbatim reproduction of any instructions and descriptions of apparatus. Especially lengthy verbal instructions and illustrations of apparatus may be attached as an appendix to the main report; however, all important excerpts and descriptions should appear in the method section proper. It should be sufficiently complete so that anyone, by reading the paper alone, could conduct the same experiment.

RESULTS

This section presents the data in quantitative form. Most of the experiments in this text are so simple that the means and significance levels can be presented as a part of the text without tables. For example, you might begin the results section for the exercise on Improvisation and Attitude Change (Exercise 5) as follows:

"Mean posttest scores for the comparison group subjects who were unexposed to any persuasive communication was 23.4. Subjects who only read the arguments had a mean posttest score of 29.4. And subjects who both read the arguments and improvised their own essay had a mean posttest score of 35.6. The read and write (or improvisation) experimental condition was significantly different from the read-only condition ($t = 2.93$; $p < .01$) and from the comparison group ($t = 6.43$; $p < .01$). The two experimental conditions were not significantly different from each other ($t = 1.42$). Read-only subjects, however, had significantly higher posttest scores than subjects in the comparison condition who received no persuasive materials ($t = 3.38$; $p < .01$). . . ."

In general, the results section should provide data to justify any conclusions. All general trends should be noted. Tables and/or figures may be used to clarify more complex data patterns. Laboratory reports may also include original data forms and comments about subject or experimenter behavior, either in the text or as appendices.

DISCUSSION

In this section you should make any comments which are suggested by the results you obtained in the exercise. Do the results confirm or contradict the

hypotheses offered in the introduction? Do your results suggest alterations in the method or the theory which led to the predictions? Can you think of any alterations that would improve the design? If your results came out as expected, what is the next study? How do the results correspond with or differ from other well-known theories or points of view? For example, the following paragraphs might appear in a discussion section for the improvisation exercise if the results were as described in the sample results section above.

"Both the read-only and improvisation subjects changed more than subjects who were not exposed to any form of persuasive communications. But in contrast to our hypotheses, the read and write (improvisation) procedures did not produce more attitude change than the read-only procedures."

"There are at least two possible explanations for this failure to confirm the hypothesis. First, an inspection of the essays that the subjects wrote reveals that they were generally of poor quality. In fact, counter to instruction, four subjects actually wrote essays that were *opposed* to lowering the voting age to 17. It is possible, then, that subjects in the read and write condition did not actually engage in any meaningful cognitive contact or improvisation. If this were the case, the cognitive contact theory would not predict greater attitude change. This procedure could be corrected by rewording the experimental instructions to stress the importance of writing good, innovative essay *in favor of* lowering the voting age to 17."

"Second, the level of choice may have been too low in this experiment to arouse any cognitive dissonance. Once subjects had agreed to participate in the experiment, they seemed to be committed to do whatever the experimenter asked. Thus the present results may be because of the fact that the experimental issues do not provide an adequate level of perceived choice. This could be corrected by changing the script and the experimenter's style of delivery in such a way that the subjects believed that they had greater choice in deciding to write or not write the essay."

SUMMARY

This section is optional. Since you began the paper with an abstract or overview, you may decide to omit a concluding discussion. However, some authors prefer to conclude a paper with a very brief summary.

REFERENCES

References in this book follow the standard American Psychological Association format (except in the case of some reprinted Research Examples). Only articles or books cited in the text of the report should be included in this section.

Americal Psychological Association, Council of Editors. *Publication Manual of the American Psychological Association.* (Rev. ed.) Washington, D.C.: APA, 1967.

The guidelines in the above discussion are based on suggestions for manuscript preparation contained in the Manual, which may be purchased by sending $1.50 to APA, 1200 Seventeenth Street, N.W., Washington, D.C. 20036.

Author index